AD_____

Award Number: DAMD17-03-2-0066

TITLE: Telemedicine Futures Symposium

PRINCIPAL INVESTIGATOR: Rashid L. Bashshur, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan Health System
Ann Arbor, Michigan 48109-0826

REPORT DATE: October 2004

TYPE OF REPORT: Final Proceedings

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2004	Final Proceedings (1 Oct 2003 - 30 Sep 2004)

4. TITLE AND SUBTITLE
Telemedicine Futures Symposium

5. FUNDING NUMBERS
DAMD17-03-2-0066

6. AUTHOR(S)
Rashid L. Bashshur, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan Health System
Ann Arbor, Michigan 48109-0826

E-Mail: bashshur@umich.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

No abstract provided.

14. SUBJECT TERMS
No subject terms provided.

15. NUMBER OF PAGES
89

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Second International Symposium
Future Directions for Telemedicine

The Purpose

The development of telemedicine/ telehealth/ ehealth (telemedicine for short) has reached a point where there is need to provide a synthesis of the evidence pertaining to clinical outcomes and costs to inform telemedicine policy and deployment, both nationally and internationally. This synthesis must assess the scientific adequacy of the available evidence and also suggest appropriate methodologies to guide future research. Further, there is a need to consider significant technological innovations likely to affect the reliability, efficiency, productivity, ubiquity, and appeal of telemedicine. Recognizing this, the Telemedicine Resource Center of the University of Michigan Health System will hold a symposium on Future Directions for Telemedicine. The symposium will be held in Ann Arbor, Michigan on May 20-22, 2004.

The Topics

In order to meet the purpose of the Symposium, five topical areas have been identified for expert analysis, presentation and discussion. These include:

* Synthesis of the evidence from telemedicine evaluation studies, both US and international, with a special focus on the scientific validity of the evidence and suggestions for future evaluation of telemedicine

* Review of current economic studies and methods for measuring the cost of telemedicine with a focus on measurement tools and analytic techniques for assessing costs and benefits as well as recommendations for pragmatic methodologies and techniques for the economic analysis of telemedicine applications in various health care settings

* Synthesis of telemedicine research findings pertaining to clinical outcomes of telemedicine with special emphasis on research design and outcome measurement in a variety of health care settings

* Description of emerging technologies, including wireless and broad band systems, for telemedicine and clinical decision support with an emphasis on the practicality, advantages and disadvantages, and potential cost-effectiveness of their application

* Description of the architecture for interoperability to support telemedicine services, the barriers for their implementation and recommendations for addressing these barriers in the United States and on a global basis

The Framework

The symposium framework includes presentation of papers on each topic in plenary sessions. Each presentation will be followed by a panel discussion and, subsequently by comments from the audience. In order to facilitate discussion, all registrants will receive copies of each paper in advance.

Participants

Participants in this symposium include academicians in medicine, public health, engineering, and information science; biomedical and health services researchers concerned with issues of biotechnology, access to care, cost and quality; health care providers in medicine and public health; and, health policymakers and program developers at the state, national and international levels. CME Credits will be provided to symposium participants (*).

(*) This activity has been planned and implemented in accordance with the Essential Areas and Policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint sponsorship of the University of Michigan Medical School, Telemedicine Resource Center and the sponsors listed below. The University of Michigan Medical School is accredited by the ACCME to provide continuing medical education for physicians and takes responsability for the content, quality, and scientific integrity of this CME activity.

The University of Michigan Medical School designates this educational activity for a maximum of 8.5 category 1 credits toward the AMA Physician's Recognition Award. Each physician should claim only those credits that he/she actually spent in educational activity.

SPONSORS

University of Michigan Medical School

Telemedicine Resource Center

National Library of Medicine

Agency for Health Care Research and Quality

Office for the Advancement of Telehealth

The U.S. Army Medical Research & Materiel Command's

(USAMRMC)

Telemedicine & Advanced Technology Research Center (TATRC)

Internet2

Altarum

CO-SPONSORS

American Telemedicine Association

European Institute of Telemedicine

International Society for Telemedicine

World Health Organization

SUPPORTERS

AMD Telemedicine

Cybernet Medical Corporation

Tentative Program

Thursday, May 20, 2004
4:00 - 7:00 p.m
Michigan League, 2th Floor Foyer
Registration.
Refreshments

Friday, May 21, 2004
8:00 - 9:00 a.m.
2th Floor Foyer
Continental Breakfast

9:00 - 9:15 a.m.
Ballroom
Welcome
Robert Kelch, M.D. Executive Vice President for Medical Affairs
Mary Sue Coleman, Ph.D. President, University of Michigan

9:15 - 10:35 a.m.
Session I:
Telemedicine Evaluation: Methodology and Findings
Chair: Mamoru Watanabe M.D, Ph.D
Presenters: Rashid Bashshur, Ph.D and Gary Shannon, Ph.D.
Discussants: Penny Jennett, Ph.D., Harry McConnell, M.D., Ronald Weinstein, M.D., Pamela Whitten, Ph.D., Peter Yellowlees, M.D.

10:35-10:50 a.m.
Health Break

10:50 a.m. - 12:20 p.m.
Session II:
Telemedicine Effects on Clinical Outcomes
Chairs: Jay Sanders, M.D., Louis Lareng, M.D
Presenters: Joseph Kvedar, M.D. and Jim Grigsby Ph.D.
Discussants: Michael Kienzle, M.D., Ron Merrell, M.D., Michele Nypaver, M.D., Eric Tangalos, M.D.

12:20 - 1:20 p.m.
Lunch
2nd Floor Foyer

1:20 - 2:50 p.m.
Session III:
Economic Analysis and Cost Measurement
Chair: Salah Mandil, Ph.D.
Presenters: Andrew Cameron, Ph.D. and Tim Reardon Ph.D.
Discussants: David Butz, Ph.D., Michael Hillman, M.D., Mark Janczewski, Col, Douglas Perednia, M.D, Dena Puskin Sc.D, Stuart Speedie, Ph.D.

2:50 - 3:10 p.m.
Health Break

3:10 - 4:40 p.m.
Session IV:
Emerging Technologies for Telemedicine and Clinical Decision Support
Chair: Pablo Pulido, M.D.
Presenters: Frank Ferrante, MSEE
Discussants: Michael Ackerman, Ph.D., Jonathan Linkous, Martha Pollack, Ph.D. Jonathan Silverstein, M.D., MS, FACS, Mark VanderWerf, Adriana Velazquez

6:00 - 8:30 p.m.
Reception and Dinner
Presenter: Fawwaz Ulaby, Ph.D., Vice President for Research, University of Michigan
Keynote Speaker: The Honorable Julio Frenk, M.D., Ph.D., Minister of Health, Mexico

---

Saturday, May 22, 2004

8:00 - 9:00 a.m.
Continental Breakfast
2nd Floor Foyer

9:00 - 10:30 a.m.
Session V:
Interoperable Architectures to Support Telemedicine Nationally and Internationally
Chair: David Forslund Ph.D.
Presenter: Rick Craft
Discussants: Charles Doarn, MBA, Pronab Ganguly, Haim Kilov, Mary Kratz, Jerry Ledlow, Ph.D., Jeff Sutherland, Ph.D.

10:30 - 11:00 a.m.
Closing Remarks

DRAFT: For Discussion Purposes Only

Not Ready for Publication or Quotation

Telemedicine Evaluation

Rashid Bashshur, Gary Shannon,

and Hasan Sapci

Introduction

To date, the major drive behind the development of telemedicine has been its intuitive appeal as an effective substitute for in-person medical care where the provider and the recipient of care are not located in the same place at the same time, and as an effective tool in developing integrated systems of care. This appeal is supported by substantial experiential data in various settings and clinical applications in the United States and numerous other countries, and more importantly by a growing body of empirical knowledge that is mostly supportive, but has yet to reach the level of conclusive and definitive evidence necessary for universal acceptance. Ironically, at this stage in its development, it is possible to reach conflicting conclusions with equally convincing justification and reason with regard to the necessity and significance of evaluation of telemedicine,

On the one hand, one can claim that the final pattern of adoption and diffusion of telemedicine, as an integral component of the medical care system, will ultimately depend on the hard evidence produced by rigorous scientific studies that evaluate its benefits and costs. On the other hand, one can also claim legitimately that attempts to evaluate telemedicine as a well-defined and discrete modality of care are, in the end, futile; and, moreover, that the underlying technology of telemedicine has developed a life of its own, which continues to evolve on a fast pace. As well, telemedicine applications continue to proliferate and now encompass all clinical areas in addition to public health and medical and health education. The futility derives from the fact that telemedicine activity is rapidly evolving and, therefore, it is difficult if not impossible to get a specific technology/human resource/application "fix", call it telemedicine in a specified version or form, and differentiate it from neighboring concepts.

Indeed, if the field is in constant flux, the most appropriate evaluation should be aimed at investigating the benefits and costs of alternative modalities and various dynamic combinations and configurations of technology, human resources, and health applications. We chose to combine elements of both arguments in this paper in order to move our thinking on telemedicine evaluation as well the field of telemedicine forward. At the same time, we must recognize the need to be creative and flexible in evaluating this field to satisfy the requirements of various stakeholders. But, in the end, the goal of

evaluation should be to produce objective evidence regarding the merits and problems of this field. In this paper, we:

- Review the status of evaluation research in telemedicine from both methodological and substantive perspectives
- Discuss basic requirements for the scientific evaluation of telemedicine, and,
- Propose appropriate strategies/methodologies for consideration that can be used in future evaluation studies and in reaching closure on the available evidence from the extant literature.

More specifically, we will describe in general terms the origin and goals of program evaluation research, the objects of evaluation in health care, a basic typology of evaluation methods, and the status of evaluation research in telemedicine and how to interpret the evidence produced to date. As a detailed illustration of how we may reach closure regarding the available evidence, we provide a theoretical derivation for evaluating the impact of telemedicine on access to care. This illustration is limited to telemedicine's effects on access since the following two other presentations will focus on outcomes and cost respectively.

As part of the status report, a synopsis of the conclusions reached by other investigators who have conducted literature reviews (a review of reviews) regarding the evaluation of this field is presented. A number of published comprehensive and painstaking literature reviews of the work done in this field is available, and will be referenced later. The results of these reviews will be utilized as the logical foundation for developing alternative strategies for future investigation and analysis of evidence. Finally, we conclude with a discussion of basic options for future evaluative research and analysis of the available empirical evidence within the context of these alternative strategies.

Basic Issues in Evaluation

It may be appreciated that the evaluation of health care programs combines scientific requirements and political realities, which are often incompatible. The scientific requirements pertain to the need for robust research designs, reliable and valid measurement, as well as rigorous methods for data collection and data analysis. The political realities stem from the priorities of public policy and funding agencies and the process of allocating research funds.

Evaluation may be focused on a specific intervention or device or it may encompass a health or other program as a whole. When encompassing a program as a whole, evaluation rests on the reasonable premise that public policy or the allocations of public resources for large scale or expensive health programs ideally are based on an objective analysis of their costs and benefits. In other words, prudent policy must be informed by factual information and evidence regarding costs and benefits of alternative programs and initiatives aimed at achieving the same goals. On the other hand, evaluation and adoption of specific medical interventions or medical devices, while based on a similar assumption, has a more limited focus, namely, that the safety and effectiveness of new medical interventions and devices must be demonstrated before they can be approved for regular professional or consumer use. Conceptually and factually, telemedicine/telehealth combines elements of both health programs as well as specific medical interventions, Individual researchers, however, are at liberty to (and frequently

do) focus on one or the other; or, less frequently, on some combination of the two. In any case, currently, there is near universal consensus that public policy with regard to future public investment in telemedicine/telehealth programs must/should be based on scientific evidence regarding their benefits and costs, as compared to the alternatives. Hence, it is time for the telemedicine community to reach real consensus regarding an optimal methodology to collect the evidence and interpret the findings

The discussion here focuses on the broader program evaluation, rather than the assessment of safety and effectiveness of specific medical interventions or devices. The latter will be addressed briefly later. The issues of interest here include the origin and goals of program evaluation, a typology of evaluation (or classification of what is evaluated), evaluation methods (or approaches), and the inherent methodological difficulties in evaluating telemedicine programs.

Origin and Goals of Program Evaluation.

The impetus for program evaluation research and the involvement of social scientists in public policy can be traced back to the 1930s when President Franklin Roosevelt's *New Deal* policies established a number of large scale social welfare and public works programs aimed at ending the Great Depression and invigorating the economy. Indeed, the development of the welfare-oriented state in the 1930s in the U.S. ushered the involvement of social science in public policy. Some of the best empirical social research during that decade was made possible by the Works Progress Administration (WPA), and the National Youth Administration (NYA), which paid for research assistants and helped unemployed academics. Nonetheless, as a method of inquiry, evaluation research did not introduce any new or different research designs or analytic techniques. Hence, it is fair to say that evaluation research does not have a unique or different methodology. Evaluators use any or all of standard methodologies, including clinical trials, sample surveys, and even the dreaded focus group which bypasses nearly all requirements for rigorous research.

Most of the early work on the evaluation of the New Deal was done by the Roosevelt Brain Trust, all of whom were economists, but no sociologists, psychologists, or anthropologists. Some thirty years later, the Great Society programs of the 1960's represented an expansion of the New Deal, reflecting the political climate of the time. They were aimed at reducing, if not eliminating, the significant economic, social, educational, and health disparities in this country. In retrospect, the War on Poverty of the 60s was established without any empirical assessment of New Deal programs or any other federally sponsored social or public work programs for that matter. Indeed, we know very little, if any thing, about the actual impact of specific New Deal programs in relation to their costs and benefits, except that unemployment rates declined and the Great Depression came to an end. In contrast, we do have some data to make reasonable assessment of the War on Poverty in terms of employment trends, health and vital statistics, enrollment in educational programs as well as statistics on crime. However, none of these assessments would rise to the level of establishing cause and effect relationships between specific interventions and outcomes. Therefore, we do not know for sure what and to what the degree specific efforts in health, civil rights, early education and others sectors or the specific cost/benefit ratios that may have accrued from the substantial national investment in these programs. What was true then and is still true today is that the politics of social program implementation did not mesh with the

scientific requirements necessary for producing accurate information. At the time, the overriding concern was to reduce basic disparities in society that brought about major social unrest as soon as possible, without a careful analysis of various options or approaches that may be pursued for reaching this goal. Similarly, telemedicine was introduced and has been developed to address overriding public and political concerns with inequities in access to care, issues pertaining to the inequitable distribution of quality of care and, of course, the ever-increasing cost of health care.

Objects of Evaluation in Health Care

As explained earlier, there are two types of evaluation research in health care: (1) the first is concerned with testing the efficacy/effectiveness and safety of specific medical interventions, devices, or medications (typically used by the Food and Drug Administration as a requisite for their approval for professional or consumer use); and (2) the second is the assessment of health program performance and achievement in terms of stated goals and objectives. Both types are complex, and both require the use of rigorous scientific methods. However, testing safety and effectiveness of specific devices or interventions is simpler and less problematic than program evaluation, and it is more amenable to conventional scientific methods. This is because the experimental variable is well defined; its effects are determinable; experimental allocation of cases is possible; and the outcomes of interest are usually observable within a limited time frame or the life time of the clinical trial.

On the other hand, program evaluation is made more complex by virtue of the fact that the experimental variable(s) are not usually clear or specific; experimental allocation is difficult at best; and, midstream/mid-course changes in the program are often encountered. Moreover, programs have a variety of effects: direct and indirect; intended and unintended; and, immediate and delayed. It is near impossible to capture all these effects within the life span of a sponsored project.

Program evaluation can serve several policy objectives, (a) to determine success or failure of programs in reaching explicit policy objectives; (b) to ascertain if there is a less costly alternative to achieve the same objectives; (c) to determine if the program has undesirable unintended effects; and/or (c) to make informed policy decisions regarding continuation, termination or change in the program, and whether to increase, maintain or decrease funding for it.

Since its findings are based on scientific evidence, evaluation research has a significant effect on normalizing new technology in the academic community, and for its acceptance among practitioners and policymakers. If rendered faithfully, it can establish the basic facts that become the prevailing wisdom. Because its methodology consists of a set of mediating and standardized practices based on the logic of the scientific method, the results are credible. Indeed, the evidence produced by rigorous science provides the "seal of approval" for professionals and the public at large. But, unlike basic research that tests specific theory or explicit hypotheses derived from theory, evaluation research is based largely on a conceptualization and measurement of operating systems in the real world, and is aimed at determining their success or failure in achieving explicit goals and objectives as well as their unintended effects.

Evaluation typology

Program evaluation categories are based largely on measurement feasibility and on the developmental stage of program implementation and development.

Evaluability Assessment can be conducted when starting or even planning large scale programs with a serious intent on having the program evaluated systematically subsequent to its implementation. This type of assessment can be used effectively to frame the research questions; to determine the research design requirements; to develop the measurement tools and the data collection schemes; and to anticipate the analytical and statistical methods to be used in manipulating the data. Its importance derives from the ability to clarify the goals and objectives of the program at the outset; to make explicit the intended effects, both short term and long term; to identify the variables of interest, and how they will be measured, and, finally to specify the analytical tools for interpreting the findings when they become available.

Documentation Evaluation is usually optional, and it is comprised of a narrative description of the actual implementation of the program. To be useful, documentation must include a faithful rendition of the steps, procedures, and protocols that were used in the implementation process, as well as a reliable assessment of missteps, pitfalls and problems encountered during implementation; and, how they were addressed. Obviously, the important rationale for documentation evaluation is to provide potential new entrants to the field with important information on successful and unsuccessful approaches, pitfalls to be avoided, and possible ways to deal with problems that arise.

Once the program is in place, data can be gathered on its effects on the process and/or outcome of care, typically referred to as **formative and outcome evaluation**. In many instances, large scale programs are based on philosophical principles that call for evaluation. For instance, telemedicine is intended to bring about fundamental changes in the care process, changes that eliminate not only geographic but also social and economic disparities in access to health care. It is expected to provide a less costly alternative to in-person care by streamlining the care process, reducing the need for travel, and by providing effective substitutions in terms of site of care and provider mix. It is also expected to assist in developing integrated systems of care as well as reducing physician isolation from mainstream medicine, thereby improving the overall quality of care. These expectations have prompted the Congress of the United States to endorse telemedicine and provide financial support for telemedicine projects, in sharp contrast to the more cautious and contemplative approach by other branches of government, such as HCFA (now CMS) for fear of potentially devastating effects of paying for pent up demand for care especially in under-served areas.

Despite the inherent interest in outcome evaluation, the vast majority of evaluation research focuses, by necessity, on process variables. **Formative or process evaluation** can capture information rich in behavioral, attitudinal and cognitive changes that are likely to occur in the short term, and can also be attributed to the program both methodologically and logically. Moreover, because these changes can occur in the short term and can be logically linked to the intervention, fewer validity threats are encountered. This makes it more feasible to rule out certain rival hypotheses or explanations for the observed trends in the data. Process variables are significant where there is a clear and logical connection between them as precursors and the ultimate outcomes. Indeed, formative evaluation is quite appropriate and useful when the variables used logically predate and foretell the expected outcomes.

Summative or Outcome Evaluation provides definitive evidence regarding the intended effects of a program, but only when implemented rigorously. The clearer, the

fewer, and the more explicit the objectives of the program the more definitive can be the findings from outcome evaluation. Subsequently, the ultimate justification of a program would then rest on the extent to which it has demonstrable positive effects or favorable cost/benefit ratios.

Since dramatic effects can be produced by the introduction of new hardware and software, there is often acute interest in "technological fixes" that are easy to install and likely to produce dramatic effects in the short term. Unfortunately, the introduction of technological fixes is also associated with high expectations, or expectations set at levels that are difficult to achieve within a short time frame. The irony here is that even moderate success may fall short of expectation. Moreover, technological fixes often have unforeseen and unintended effects, both positive and negative. These are usually difficult to account for and to measure.

The most notable model in the use of summative evaluation for rational decision or policy making is cost-benefit analysis (CBA). Alas, CBA has the inherent limitation of converting all benefits and costs into monetary values, including such intangibles as the monetary value of duration of life, quality of life, and convenience. However, if its assumptions and metric are accepted, cost-benefit analysis can provide a rational basis for answering questions regarding critical choices among alternatives which may have varied and profound effects, including unanticipated effects. This topic is discussed in more detail in the next session dealing with economic analysis.

Summative evaluation, as one might expect, is of particular importance to policy and decision makers who are inherently interested in the ultimate results of programs to justify public expenditures and resource allocation decisions at the national, regional, and institutional levels.

Status of Telemedicine Evaluation

The second generation of telemedicine activity beginning in the early 1990s in the United States and several other countries provided substantial funding for research and development. But, by virtue of the political process in the United States and perhaps by necessity, the bulk of the funding was allocated to program development. Of course, programs had to be planned, organized, staffed, and launched. More specifically, infrastructure had to be installed, tested, and refined. Providers and patients had to be recruited and trained. And, finally operational procedures, protocols and logistics had to be developed and routinized. In other words, the initial requirements for establishing telemedicine programs are quite extensive and time consuming. Nonetheless, at the same time they were establishing themselves, most funding agencies – perhaps as they should - required them to conduct an evaluation of their programs within the two or three year span of the funding cycle often dealing with complex issues such as cost effectiveness and clinical outcomes during limited time frames.

The political and geographic considerations in Federal funding for telemedicine programs in the United States may have inadvertently preempted the opportunity to design and execute large scale randomized clinical trials to evaluate the emerging practice of telemedicine. Purposefully or not, we are left with a large number of programs widely distributed around the country, but none of them having the resources or long term support necessary for large scale clinical trials. Moreover, the projects that attempted to collect uniform data sets across several sites have yet to publish results beyond descriptive trends. In this field and perhaps others like it, there has always been a

tension between the ideal world of evaluation and the real world of program development.

Despite, or perhaps due to, a history of several decades, we have yet to agree on a uniform and precise definition of telemedicine. This has been exacerbated by a liberal trend in the nomenclature to include more inclusive terms such as telehealth and ehealth without a clear consensus on the delineation of boundaries between them and telemedicine. A uniform agreement on the content and boundaries of telemedicine and other related terms would seem to be a prerequisite for valid evaluation of this field. Without such agreement, evaluation will remain largely project-specific and, therefore, of limited generality. Today, what we call telemedicine or telehealth encompasses any existing or proposed configuration of technology, organization, and human resources; as well as single or multiple clinical, educational, and public health applications. In this regard, no single or standard referent exists as to what telemedicine does or does not represent. The lack of precise definition compromises our ability to evaluate the true concept in full fidelity, strength and integrity of the intervention (Sechrest, 1979). Accordingly, whatever definitional imperfections may exist in the configuration of a particular system under investigation will likely mask or mitigate the true effects of the concept telemedicine in its optimal form.

In summary, a review of the telemedicine research literature to date reveals that the empirical research is not based on a precise and uniform set of common parameters. Hence, this body of knowledge, though substantial in volume, is segmental and inconclusive. It is segmental in the sense that much of it focuses on specific applications rather than integrated systems of care revolving around patient care and provider and client education. Consequently, much of this information has little to do with systemic effects of telemedicine.

Furthermore, even well designed experimental studies can be misleading. For instance, diagnostic accuracy has been typically determined on the basis of agreement between in-person care (as the gold standard) and care provided by telemedicine. As Koran (1975) pointed out, however, physicians tend to disagree with each other "once in ten cases, and often...one in five cases, whether they were eliciting physical signs, interpreting roentgenograms, electrocardiograms, or electroencephalograms, making a diagnosis, recommending treatment, or evaluating the quality of care." Hence, reliability based upon "agreement" should not be equated with accuracy.

It is interesting to note that, in an attempt to control inter-observer variability, some researchers have limited the number of observers to one or two who would render the two kinds of observations, telemedicine and in-person, on the same subjects. This creates a larger problem in terms of the lack of independence in the observed data. Without independence, it is futile to conduct any meaningful statistical analysis or to test any hypotheses.

After years and scores of attempts to develop and evaluate telemedicine, the complaint continues to be found in the literature that the quality and methodologies of telemedicine evaluation studies is poor. As early as 1980 Bashshur (1980) lamented the fact that in the late 1960s and early 1970s, there had not been a broad range systematic effort to address all the basic pertinent issues in the assessment of telemedicine and, further, "policy decisions regarding the funding of telemedicine.....utilized a mixture of anecdotal and inconclusive empirical findings." In the mid-1990s, the National Library

of Medicine recognized a national need for robust telemedicine evaluation and called for the Institute of Medicine to develop a framework and guidelines for telemedicine evaluation. (IOM, 1996, Lewin 2000).

DeChant et al (1996) suggested a "staged approach" to deal with "the great variety in telemedicine applications and ...new information systems for health care delivery (that) pose(d) challenges to traditional methods of technology assessment." A systematic review of over 1,000 articles evaluating telemedicine cost-effectiveness revealed "only a few" controlled studies and the vast majority did not produce the requisite scientific evidence. (Roine et al. 2001) A review and critique literature pertaining to evaluation of telehealth "solutions" called for consistency in the nature of evaluation activities "-- if achievable -- " and also suggested that this would certainly not occur in the short term. (Health Canada 2000) The Canadian report concluded that "most telehealth applications are either not well evaluated, or are evaluated in an ad hoc manner." A review of clinical outcomes in telemedicine concluded that in only a small number of studies does evidence of its benefits exist, and there is need for further randomized clinical trials (Hersh et al. 2001).

A 2001 Report to Congress on Telemedicine indicated a lack of statistical evidence in most of the studies examined and expressed "hope" for more statistically robust studies in the near future. (Congress 2001) In 2002, an editorial in the Journal of the American Medical Informatics Association suggested that "the (telehealth) literature does not contain an adequate evaluation of telemedicine despite years of application of the technology and several calls to action." (Stead 2002) Also, a literature review of telemedicine cost-effectiveness and patient satisfaction concluded that for the former "there is no good evidence that telemedicine is [or is not] a cost effective means of delivering health care" (parentheses ours) and, for the latter, methodological deficiencies (small sample sizes, context, and study designs) limited the generality of the findings. (Whitten et al. 2002; Mair and Whitten, 2000).

Another review reported the scarcity of good quality studies providing the evidence for the benefits of telemedicine. (Hailey, Roine and Ohinmaa, 2002) A review of clinical and educational telepsychiatry applications did not incorporate cost studies because the quality of data in the literature was "suboptimal and little ....has been collected in a systematic, controlled prospective fashion." (Hilty et al., 2004) And, "more short- and long-term quantitative and qualitative research is warranted on clinical outcomes, predictors of satisfaction, costs, and educational outcomes." Drake (2003) raises the basic question regarding the future in his literature review of telemedicine evaluation in 2003, namely: (1) "Even when mature evaluation techniques are extrapolated to telemedicine, after the studies are published, the most important question appears to remain: Is telemedicine technology worth advancing, and how as a society do we decide?" (Drake 2003)

These studies and conclusions are representative of recent systematic reviews of telemedicine evaluation studies. We may draw the general conclusion, therefore, that with few exceptions (e.g., Shea et al. 2002) the research in this field has yet to produce an adequate body of empirical findings that rises to the level of conclusive evidence as traditionally defined. And, therefore, in a strict sense, we cannot ascertain with a programmatic degree of certainty the precise or specific effects of telemedicine on

access, cost and quality or the interaction between these effects, which are important issues from a public policy standpoint.

Nevertheless, the bulk of the research evidence to date has demonstrated the feasibility of telemedicine in almost all clinical and diagnostic applications. More than a decade ago, Grigsby and associates (1993) reported that teleradiology is "effective across all specialties." Indeed, the question of feasibility of telemedicine in all other clinical applications has been put to rest for quite some time, albeit at different levels of confidence in the different applications. Interestingly, clinical applications have been classified by the level of maturity on the basis of several performance measures and attributes, including volume and quality of research findings, demonstrable technical feasibility, diagnostic accuracy, specificity, clinical outcome and cost effectiveness (Krupinski, et al 2001). According to this classification, teleradiology and telepathology occupy the first tier, and are labeled as "mature" applications. The second tier, or "maturing" applications consists of telepsychiatry, teledermatology, telecardiology, and teleopthalmology. All other applications are in the third tier, and labeled as "emerging" applications. Of interest here is the fact that all research findings pertaining to clinical effectiveness across a variety of applications and at various levels of maturation are linked to the specific technology that was used in the application, and all technical failures and limitations are totally attributable to specific sets of technologies that were used, and have been improved since. Hence, no inherent limitations have been reported.

Finally, upon careful reflection, several observations can be made regarding the extant evaluation literature.

- None of the systems or programs that were evaluated in the United States or other countries had fully exploited the technological capabilities at their disposal. Indeed, for a variety of reasons, all programs have substantially underutilized the capabilities of the technology at their disposal. Hence, any economic assessment that takes capital cost into account that does not incorporate sensitivity analysis under various assumptions of utilization volume, resource capacity, maturation, or steady state operation is likely to underestimate the potential return on investment. Since capital cost is fixed, it is obvious that the more and the more efficient the program is used the more favorable would be the return on investment. Moreover, given conditions of under-use, the logical question continues to be whether simpler and lower-cost technology (i.e., lower investment) could be more fully, and therefore, more cost-effectively utilized. But even this argument may have flaws because scale is important, and some minimal level of investment and size operation may be necessary.
- In addition to the above, nearly all telemedicine projects have missed significant opportunities to utilize the enormous capabilities of information technology to create fully integrated health care delivery systems. It may be appreciated that the underlying technology provides not only ready connectivity between various parts of the delivery system, but perhaps more importantly ready access to information from various sources within and outside the institution for various components of patient care, protocols-driven disease management programs, and clinical decision support. The effective use of these rich capabilities and their

integration into a single organization or system should reduce so called errors, enhance efficiency, minimize unnecessary duplication and waste, and streamline the care process.

Nearly all projects had narrowly defined clinical or educational functions and specific target populations, and they tended to limit themselves to these functions. This left unattended important questions pertaining to the full potential of the information technology at their disposal. This is akin to having a desk top computer at the office, but using it only for electronic mail. While it makes perfect sense to use the computer for mail only, one should not lose sight of the other numerous capabilities and the missed opportunities when they are not fully exploited. The ramifications on cost/benefit ratios are quite clear.

- Given these observed limitations in the scope of the applications to date, it is not surprising that we have yet to assess adequately the broader systemic effects of telemedicine on use of service, referral and admission rates, access to care, quality and cost of care in a way that allows true generalization about these effects. As well we have yet to achieve an in-depth understanding of the tradeoffs and interactions between these various effects as discussed elsewhere in this paper..

Evaluation Research - Proposed Approaches

The proposed approaches for valid evaluation of telemedicine are discussed here in terms of (1) prerequisites for optimal evaluation; (2) research design requirements; and 3) analytic tools for causal inference.

Prerequisites for Optimal Evaluation

Past evaluation of telemedicine programs was significantly hampered by the absence of optimally operating systems that had progressed sufficiently on the learning curve to permit a valid evaluation under optimal or at least steady state conditions. Evaluating programs under these circumstances runs the risk of simply reflecting imperfections in design or implementation rather than inherent flaws in the basic concept of harnessing information technology for the remote delivery of care or a necessary failure in the capability of the intervention. Furthermore, even when administered in optimal fashion, it is necessary to separate the contextual effects from the experimental effects. To address this problem, Campbell (1969) proposed the "climax model" which calls for evaluating programs only after they have achieved a steady state of operation and also contained all the intended core elements and capabilities of the program. The use of this model assures that programs would have sufficient fidelity in terms of the strength, volume, and stability of programs as a precondition for assessing their true effects.

At the same time, it is clear that telemedicine constitutes an innovation "bundle" rather than a single categorical intervention. When the concept of "innovation bundle" was introduced (Rogers and Shoemaker, 1971), a precise definition was not offered. Nonetheless, an innovation bundle is expected to produce one kind of effect for the entire bundle that is typically incremental in nature. However, the telemedicine bundle is likely to produce a variety of effects that are associated with various subsets of variables. These effects may represent tradeoffs rather than being unidirectional and additive. For

instance, telemedicine presents clear transportation/communication tradeoffs, which may increase access and enhance quality while reducing opportunity cost. But, we have yet to explain or understand what happens to cost when accessibility is improved. For example, would there be a commensurate increase in demand, even pent-up demand, and a greater risk of "moral hazard," or the tendency to use medical services more because some barriers to use have declined? Would the availability of telemedicine lead to an increase in provider-induced demand on the part of those seeking to enhance their revenues from the use of the technology? Would telemedicine displace the scarce medical resources available in a highly constrained rural environment, thereby inadvertently diminish rather than enhance the available resources? Would telemedicine likely produce the paradoxical effect of promoting quality of care in underserved rural areas only to reduce reliance on telemedicine as time goes on. In other words, would its success lead to its ultimate demise? These and other important questions have yet to be addressed in definitive and comprehensive ways. Indeed, we have yet to understand whether success or failure in a specific clinical application or subsets of applications are generalizable to other applications or the entire system of care; for instance, whether success in teleradiology is indicative of success in other clinical applications or of entire programs that offer a variety of clinical services, or alternatively whether telemedicine systems are more than the sum of their parts.

While the telemedicine "package" may be innovative, all the components that comprise the bundle may not be so. In telemedicine, for example, certain traditional clinical and communication practices are likely to be maintained and "mixed in" with the new mode. As well, the broader medical care environment where these systems are installed may not alter established practice patterns simply by virtue of adding telemedicine to its established repertoire of services. In other words, neither providers nor clients may alter completely all their traditional behaviors simply by virtue of having telemedicine available.

Hence, the evaluation process has to be sufficiently refined and detailed to enable a valid analysis of the specific effects of the "newer" components as well as combinations of new and old ones, and to be able to isolate telemedicine effects independently from contextual effects. This will not be easy to do, but it is particularly salient in assessing telemedicine's effects on cost, quality and access.

While the calls for health care reform in the United States and elsewhere would ebb and flow from time to time, there is near universal and omnipresent agreement on three related national goals. Access to care should be improved, if not assured, to all people in need of care. Cost should be contained as much as possible without compromising people's health or safety. And, quality should be maintained, if not improved. With these goals in mind, rural populations are prime targets as potential beneficiaries of telemedicine because of their relative isolation, geographic dispersion, and limited economic resources.

Proponents have argued that telemedicine would accomplish all three goals effectively and efficiently not only for rural and other isolated populations but also to the home-bound and the institutionalized, as well as urban and poorly served populations. However, some critics are concerned that telemedicine may produce adverse effects by decreasing accessibility (due to displacing rural providers); increasing total cost (due to consumer pent up demand and potential provider-induced demand); and compromising

quality (by relying on a technology that may be a poor substitute for in-person care). Interestingly, when juxtaposed, the hypotheses of the advocates and the critics are enantiomorphic, that is, they mirror one another, and thereby demonstrate their centrality and the need to subject them to rigorous scientific testing.

Research Design

Two issues must be made clear in considering the choice of appropriate research designs for telemedicine evaluation. First, the controlled experiment is the ideal design for evaluation research because it calls for the random assignment of test subjects to experimental and control groups, blinded administration of the experimental variable, and pre and post measurement. No doubt, the controlled experiment is a powerful tool that would yield objective and valid results from which no one can hide. Weaker methods would yield equivocal results that can be manipulated by program managers and decision makers to suit their particularistic agendas. While such methods don't necessarily invite mischief, they provide a fig leaf for those intent on mischief. Hence, other things equal, the controlled experiment is the preferred method for testing explicit hypotheses, for ruling out rival hypotheses or explanations for observed phenomena, and for establishing cause-effect relationships between the specified variables of interest. When one or more requisites of a controlled experiment cannot be obtained in a given situation, substitute quasi-experimental designs can be used, which mimic the controlled experiment to the extent feasible. This is true for simple medical interventions, such new devices, new medications, or new procedures, as well as medical programs, such as broad range neighborhood health centers, regional medical programs and, of course, telemedicine. In brief, the essential feature of a controlled experiment, or randomized clinical trial, is the active participation of the experimenter in randomizing subjects, administering the treatment or program intervention, and measuring the variables of interest in both experimental and control group before and after the intervention.

Nonetheless, as we shall discuss later, randomized clinical trials may not be feasible or optimal in this field.

Secondly, it is not the absence of knowledge or methodological sophistication that is hampering progress in this field. Knowledge exists, and the requisite methodological tools are already in hand. They do not have to be invented. Indeed, factorial designs enable the detection of differential effects of various combinations of variables, including characteristics of clients or individuals using service (such old versus young, male versus female) various levels in severity of illness, provider mix (such as generalist versus specialist, specialty, experience, non-MD providers.) concentrating on features that either singly or in combination produce the greatest effects. However, in the real world, it is likely impracticable, in some situations impossible, to conduct controlled experiments to evaluate broad range programs such as telemedicine. Among other things, the requirements of scientific inquiry on the one hand, and the political, legal and ethical imperatives on the other are not consistent. As they should, laws and regulations governing individual autonomy through informed consent and the proscription of harmful interventions are appropriate in a civilized society. However, some of these regulations can create a researcher's nightmare. Human subjects have the right to refuse participation in any study that either observes them directly, collect information about them or use information already available on them. They have the right to withdraw their consent at any time during or after the conclusion of a study; and, in certain instances, to jump

experimental allocation. In some instances, experimentation with human subjects has conjured up images of a Frankenstein monster. And yet, in other situations, the intervention may be deemed so beneficial, it may be considered unethical to withhold it from those in need of it. It is not uncommon for researcher to have to nullify or diminish the effects of the experimental variable by promising remediation to members of the control group or holding them harmless if the health plan they were assigned resulted in higher out-of-pocket cost, as occurred in the Rand Insurance Study (reference).

That serious issue aside, random assignment of patients to experimental and control groups is not possible when the number of telemedicine encounters is extremely limited. Further, situational or contextual factors cannot be held constant to rule out rival hypotheses, given the empirical diversity of programs produced by local variations in administration, variations in clinical applications and in organizational and staffing configurations, and the never ending change in the underlying technology.

There are several other impediments to the use of experimental designs in telemedicine research. (1) First, we have yet to agree on the precise scope of the program to be evaluated and to differentiate between the overall goals of the program and those of the specific applications within it. Telemedicine programs can range from single specialty service within a single delivery system such as teleradiolology to a multi-specialty multi-site network that offers a full range of medical and diagnostic services. (2) Second, as mentioned before, the prevailing laws and regulations covering the use of human subjects often challenge randomization rules, which blindly assigns subjects to experimental or control group, do not permit subjects to "jump" or transfer from one experimental allocation to another, and do not allow administrators or project managers to make midstream changes in program intervention – especially in terms of key variables that may alter the outcomes. Such changes include technological configurations (in terms of hardware and software), human resources and staffing or the organizational structure of the program.

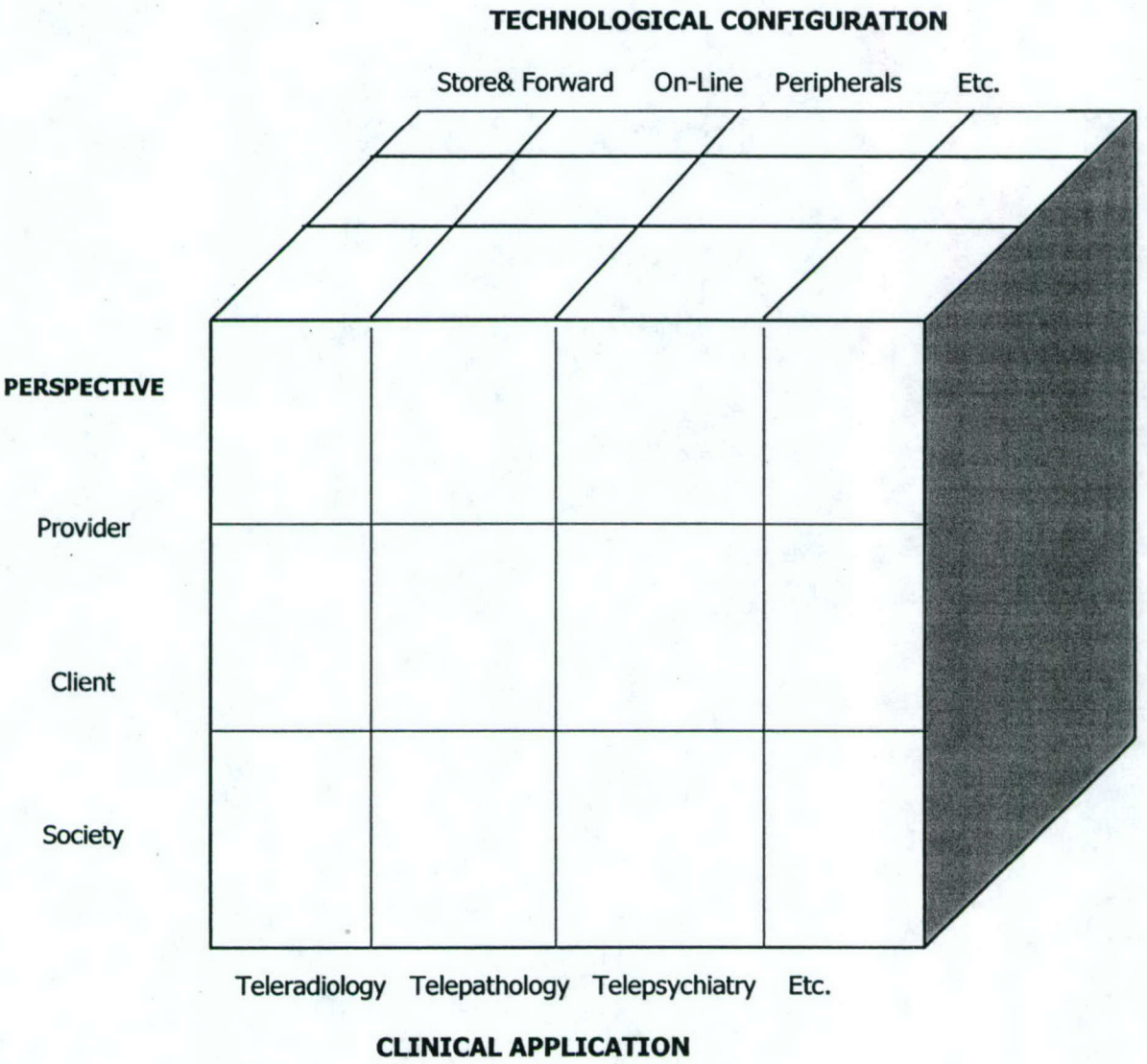
Beyond these serious difficulties, the evaluation of telemedicine faces some unique problems in terms of the nature of the experimental variable, or the unit of analysis. The usual units of analysis as well as units of observation that have been reported in the telemedicine literature typically consist of individual medical visits or encounters. But, the more appropriate units of analysis in telemedicine research may consist of episodes of care or illness episodes to be used in comparing telemedicine to in-person care. Indeed, more useful and valid information about the effects of telemedicine on utilization and cost would be captured when using episodes of care or episodes of illness rather than visits or encounters. This is because episodes can capture more information for assessing the contributions of telemedicine. Episodes can be compared between the two modalities of care, telemedicine and in-person, while controlling for severity of illness. They can provide comparative information regarding the number and duration of visits, intensity of care, substitutions (between sites and providers), timeliness of care, waiting time, and outcomes. This would enable us to ascertain whether or not telemedicine has a significant effect on the process and outcome of care. It would, for example, reveal, among other things, whether or not the use of telemedicine resulted in more timely diagnosis, shorter duration of an illness episode, and ultimately more successful outcome.

As many observers have pointed out, the concept of telemedicine continues to evolve and change in scope, application, and the underlying technology. Hence, it would be futile to evaluate it as a fixed entity. It can represent different things to different people at different times. Nonetheless, for policy purposes, evaluation may have to focus on the general attributes of the practice of telemedicine, if such parameters exist or telemedicine as a whole if it is a single entity, in order to assess its overall merit and how it best fits into the health system.

Finally, the gains or losses that might be accrued from telemedicine are likely to vary by the intended target or user, the nature of the application, and the specific technological configuration that is used in the program. More specifically, the benefits and costs of telemedicine should be viewed from the varied perspectives of providers, clients, and society at large. Each has concerns that must be identified and addressed. It should also be designed to account specific applications where it is practiced, such as teleradiology, telepsychiatry, teledermatology, and so on. Finally, it must consider the specific technological design and configuration in use, whether for example it is a synchronous or asynchronous system, bandwidth, peripheral devices and so on. This multifaceted approach can be represented by a three dimensional model adapted from Donabedian (Year), as shown in Figure 3.

Figure 3.

Three Dimension Model for Telemedicine Evaluation



The first dimension in the model represents the various clinical applications that encompass the various specialty areas of medical practice, as well as medical education. If we take the nomenclature of telehealth seriously we must add public health applications, including disease surveillance, epidemiology, health education and health behavior and health administration. The second dimension represents various technological systems and configurations in use, including synchronous online systems, asynchronous store-and-forward systems, various modalities of transmission (wired, wireless), bandwidth, and peripheral devices for diagnosis and treatment. The third dimension has to do with the specific perspectives from which an evaluation can be conducted, specifically those of clients, providers, and society at large. Obviously, despite some commonalities between them, the concerns of clients (or consumers) might differ from those of providers or of society at large. Each perspective must be measured separately. For example, on the question of cost, clients may be interested in out of pocket cost, whereas providers may be interested in revenue and return on investment. Societal interest may relate to equitable distribution of health care resources and serving unmet needs of the population. Thus, even when concerned with the same basic issues, providers, clients, and society may approach them differently because their stakes are different.

This model provides a systematic structure for the integration of research findings from different empirical studies, while leaving researchers to pursue their own interests in addressing a variety of research questions in their respective domains, to do so in an appropriate context that can combine their findings with those of others into a coherent set. Indeed, the use of the model allows for choice, flexibility, and continuity in framing the research questions and in compiling the empirical evidence, while at the same time providing the requisite structure for accumulating evidence from various studies, performed by different investigators at various times. The individual compartments or cubes within the matrix can be used singly or in various combinations. For instance, one researcher may choose to investigate the cost effectiveness of teleradiology, using store-and-forward technology on the quality of diagnostic images from a provider perspective. Another researcher may assess the systemic effects of a telemedicine program on access, cost, and quality of care from a societal perspective. The findings from each study or set of studies on the same topic may be used to fill in one compartment or cube in the model, which can be subsequently added to findings from other studies for a cumulative total that combines all other studies in the other compartments until the picture is complete. In the end, this cumulative process would answer the pressing questions in the field, and the ultimate impact would be enormous.

The derivation of the research questions for evaluation constitutes yet another fourth dimension in this model, which is concerned with the intended and unintended effects of the program, both short term and long term, and possibly its technical performance. The research questions stem from the goals of the evaluation, or what we re trying to learn.

Typology of Research Questions

Viewed broadly, there are essentially there are two types of research questions that are appropriate for telemedicine evaluation. The first derives from biomedical or clinical research, and it encompasses issues of clinical effectiveness/efficacy and safety. More specifically, biomedical research seeks to ascertain the accuracy, precision,

reliability, and sensitivity as well as safety (or side effects) of specific technological components in providing diagnostic information. The basic concern of biomedical research is to determine the extent to which specified technological components meet or exceed clinical standards of performance as compared to in-person observation and measurement. The second type or set of research questions derives from health services research. Here, the questions are typically posed after the conclusion of biomedical or clinical research, and the technology is certified as safe and effective. Indeed, in a perfectly logical world, health services research would begin when biomedical research ends, and it would focus on the effects of telemedicine on various aspects of health care delivery, namely, access, cost and quality.

Whereas biomedical research is indispensable and must be continued, it seems to be faced with overwhelming odds against proving anything definitive on a permanent basis in relation to the true merits of the concept of telemedicine. This is because technological developments tend to obviate observed imperfections in the technology, and the field continues in its press to find effective substitutes for the five senses. The technological horizon keeps expanding. In turn, these technological improvements have a direct effect on cost/benefit ratios and the tradeoff between cost and quality. In brief, technological advances tend to obviate limitations attributed to specific technologies in diagnostic accuracy and reliability. Moreover, the ongoing reductions render the observed cost/benefit ratios obsolete.

The two types of research methodologies and pertinent questions are shown in Figure 4, together with some proposed measures.

Figure 4
Types of Research in Telemedicine, Research Questions, and Measures

	Research Questions	Measures
Biomedical Research	Safety, Effectiveness	Precision, Accuracy, Specificity, Reliability
Health Services Health care	Access, Quality, Cost	Utilization, Referral, Convenience, Opportunity Cost, etc

Questions and hypotheses stemming from both biomedical and health services research can be readily translated into specific measures to be used in data collection and subsequent analysis. Some of these measures are simple and straightforward, and some are not. All of them require validity and reliability testing to ensure that they actually measure what they purport to measure in a consistent fashion. The telemedicine literature

already contains several standardized measures borrowed from social science research, which may be used by researchers in this field.

Telemedicine Evaluation: Revised Strategy

The initial argument here was for development of large scale research programs in order to achieve the requisite scientific rigor for evaluating each dimension of telemedicine, within a proposed three-dimensional model. Nevertheless, the lack of potential funding for large scale clinical trials, logistical and other constraints described earlier lead us to offer an alternative that can be pursued immediately; namely, to explore the use of a non-traditional evaluation strategy to assess the impact of telemedicine programs, based on existing and new data as they become available.

In this section, following a brief statement pertaining to the contextual framework of telemedicine development and its expected role in improving access, we focus first on a question fundamental to evaluating the association between telemedicine and the impact on access to care, namely, the notion of cause and effect and its measurement. Finally, we propose and discuss a new strategy for assessing the impact of telemedicine on access, quality, and cost of medical care. But, for illustrative purposes and because issues of cost and quality will be considered in separate papers, we limit our consideration here to the association between telemedicine and access to care.

Contextual Background

Concern for the inequitable geographic distribution of medical care was expressed as early as the mid-19th century when the (then) nascent American Medical Association charged state medical societies to assess the geographic availability of medical care. Several states reported that physicians were fewer and less likely to have adequate medical training in rural areas when compared to urban. Development and diffusion of advanced technologies such as the automobile and telephone in the late 19th and early 20th centuries reshaped the argument regarding the geographic access to care. But, the emergence of specialized medicine and office/clinic/hospital-based medical care in the mid–20th Century exacerbated the geographical disparity in the distribution of medical care resources between rural and urban areas, as well as between developed and developing countries.

Historically in the United States, attempts to redress the inequitable geographic distribution of medical care such as the Sears Roebuck Foundation's Community Medical Assistance Program (date), the Hospital Survey and Construction (Hill-Burton) Act (1946), and the National Health Service Corps created as part of the U. S. Public Health Service under the Emergency Health Personnel Act (1970) are among the leading initiatives aimed at redressing the prevailing geographic disparities in availability of health care resources. But all were met with mixed results and generally limited success. Thus, the excitement that met the introduction of telemedicine initially in the early 1970s and now, was based, in large part, on past failures to redistribute medical care resources physically and the emerging virtual distribution which would provide underserved populations with ready access to medical care when needed. Virtual distribution of health resources is expected to diminish opportunity cost for clients or patients, and it should enhance quality of care, and hence improve health status. Other factors contributing to the excitement include the potential for creating integrated medical systems, and reduction in so-called medical errors.

Causal Inference

Any evaluation of the impact of telemedicine on access implicitly or explicitly assumes a cause and effect relationship. For philosophers, the notion of cause and effect is murky at best. Hume and others argued, for example, that we can never know with certainty whether two events are causally related. There is simply a constant temporal conjunction observed repeatedly, and, "therefore what we term 'causality' can be nothing other than mere constant conjunction of the idea of the cause with that of the effect," and, this is known from experience. (Fowler 1998) But we are living in the real world, not that of philosophers, and we need to know *to our satisfaction* whether there is a *causal* association between telemedicine and access and, moreover, the direction and magnitude of that association.

Here, we may be informed by a similar, perhaps simpler, situation that occurred in the development of epidemiology, namely, the evolution of criteria that would satisfy researchers that there was indeed "causality" between a disease agent (virus, bacterium, parasite, etc.) and a disease. Initially, following the discoveries of Pasteur in the late 19th Century, the so-called Henle-Koch "postulates" were used. These postulates stated that an agent is a cause of a disease if it is present in all the affected persons ("necessary" cause), it is absent in healthy subjects ("sufficient" cause), and it can be inoculated into an animal to induce the same diseases that it causes in humans. However, implementation of these criteria in epidemiology proved impracticable, if not impossible.

Later, Sir Austin Bradford Hill's (1965) proposed nine criteria of causation as minimal conditions to establish a causal relationship between two events. Nonetheless, he cautioned that: "None of these nine viewpoints can bring indisputable evidence for or against a cause and effect hypothesis…What they can do, with greater or less strength, is to help answer the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?" He even went so far as to suggest that "none of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis, and none can be considered as a *sine qua non*." However, Rothman (1996) points out if the "cause" does not precede the effect, that indeed the temporality criterion is indisputable evidence that the association is not causal.

While the criteria established by Hill (and elaborated by others) were developed as a research tool in epidemiology, they may be equally applicable in other social sciences seeking to establish causal relationships among social phenomena, especially when it is not feasible to conduct prospective large scale randomized clinical trials. It must be remembered, however, that Hill's criteria and, historically, discussions on causality, have proceeded once a statistically significant relationship between a potential causal factor and a disease has been found. In at least one instance, this threshold was ignored for apparent political purposes by a government agency leading to considerable criticism from the scientific community pertaining to the role of environmental carcinogens in a "cancer epidemic."(Milloy and Gough 2004) We have adopted/adapted Hill's criteria for proposing a new strategy in the evaluation of telemedicine.

Armed with these criteria, epidemiologists have been able to investigate suspected single- and multi-factor causal relationships and to advise the public and private health sectors on health policy decisions. Essentially, the process is one of "triangulation," or

the application and combination of several research methodologies in the study of the same phenomenon.

By combining multiple sources of data, theories, methods, and empirical materials, triangulation may overcome the weakness of research design or intrinsic biases that come from small studies with inconclusive findings. Researchers are able to "triangulate" or draw significant conclusions in terms of the convergence of findings from various studies and methodological approaches to the same topic.

Addressing the field of Human-Computer Interaction, Mackay (2004) suggests that to mature as a field, there must be a shared research context and the need for triangulation: "using different techniques to operationalize behavior while attempting to measure the same phenomenon... which greatly increase the generality and construct validity of the findings." A study of air-traffic controllers, for example, included biological analysis of sleep patterns, laboratory experiments of different user interface strategies, computer simulations in the use of new tools used by controllers, cognitive models of air traffic control's activities and ethnographic studies of controllers at work.(Mackay and Fayard (1997)

Triangulating among various forms and sources of data means that if an observation holds across a variety of contexts viewed from different research perspectives, then we are assured of the integrity of the finding. It has also been suggested that through the triangulation of qualitative and quantitative data, a "more complete, holistic, and contextual portrayal of the research issue is captured." (Jick 1979; Mitchell 1986; Handfeld et al. 1996) A number of triangulation types have been identified:

- **Data Triangulation**: The use of multiple data sources to help understand a phenomenon
- **Methods Triangulation**: The use of multiple research methods to study a phenomenon
- **Investigator Triangulation**: The use of multiple investigators in collecting, analyzing, and interpreting the data
- **Theory Triangulation**: The use of multiple theories and perspectives to help interpret and explain the data

While one might be tempted to downplay or even disregard a conclusion from a single data source, a conclusion derived from triangulation from multiple sources is more credible, and can be valuable and meaningful. Triangulation of information can allow even statistically limited or weak conclusions drawn from one source to be compared to conclusions drawn from other sources. In the following section, we attempt to demonstrate the use of theoretical triangulation as it applies to assessing the impact of telemedicine on access.

Evaluating Telemedicine and Access: A New Strategy

Generally, access refers to the ability of patients to use appropriate health resources in a timely manner. The concept summarizes a set of specific dimensions that describe the "fit" between a patient and the health care system (Donabedian). The Institute of Medicine (1996) suggested that access can be enhanced by increased availability of health information.

The realization of the need and appeals for scientific evaluation of the impact of telemedicine on access is longstanding. Nevertheless, the appraisal of telemedicine

access-evaluation reveals few studies that provide definitive conclusions. As is the case for evaluating cost, quality and other aspects to telemedicine, valid reasons have been offered for this seeming inability to evaluate the association between telemedicine and access adequately. These include the lack of mature telemedicine programs due largely to the recency of the innovation, as well as limited funding duration in some instances. Also complicating matters is an inherent lack of stability and the changing nature of telemedicine reflecting inclusion of innovative technology as it becomes available and applications in new areas. Additionally, to date there have been few if any large scale telemedicine experiments specifically designed and funded to conduct large scale randomized clinical trials to determine the various effects of telemedicine on cost, quality, and access and interactions among these effects.

Given these conditions, telemedicine may not lend itself to traditional methods for demonstrating cause and effect. Consequently, it is time to re-evaluate the concept and strategy of evaluation in this field.

A triangulation process may be the appropriate tool for interpreting the extant evidence. The "triangulation process" refers to conclusions from a number of sources within each criterion of causal inference. In turn, the total results from these triangulations produce a valid basis for reaching closure on the intended effect. The components of the proposed triangulation process are illustrated in Figure 5 (Causality in Telemedicine).

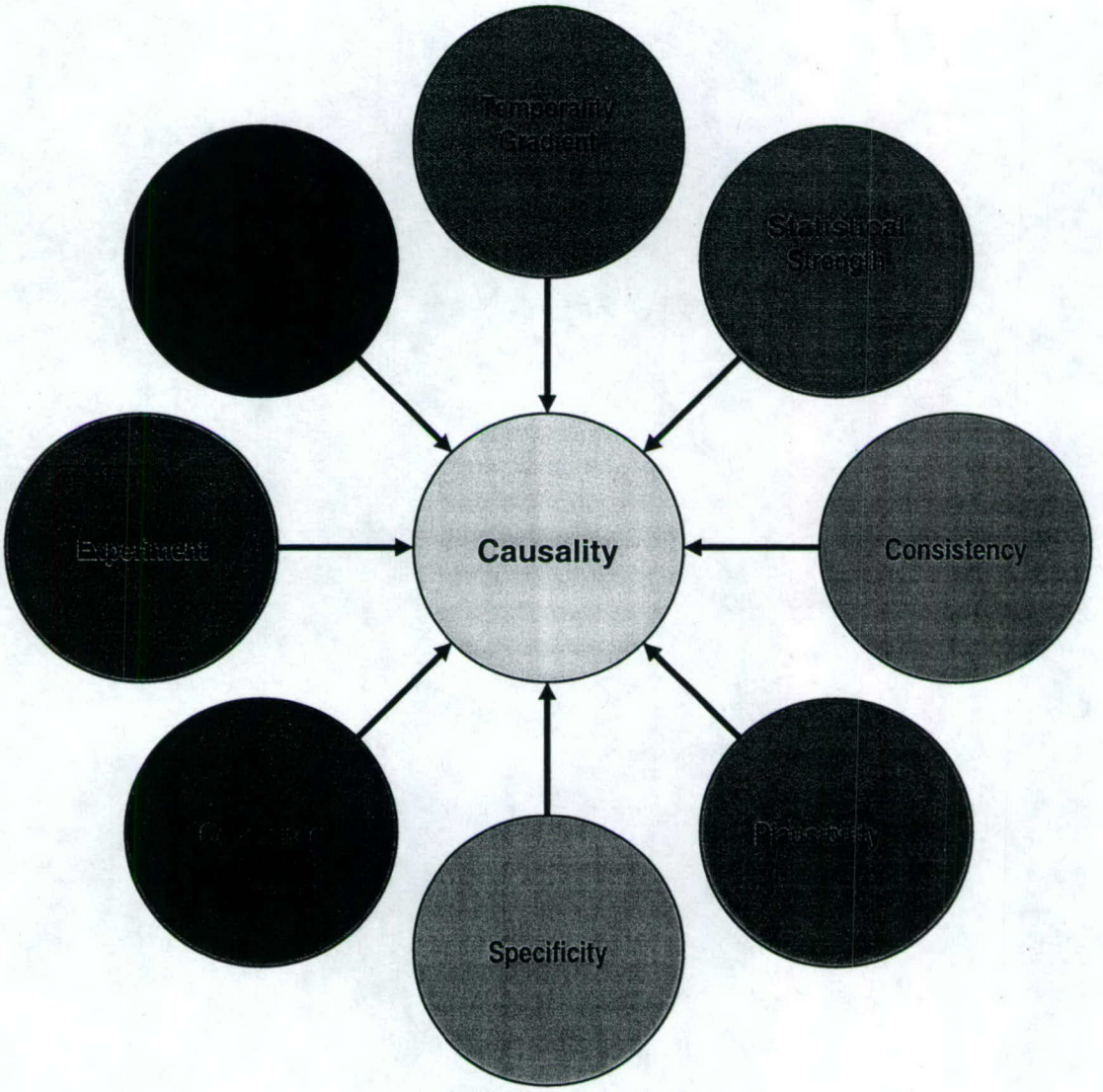


Figure 5
Causality in Telemedicine

The criteria in Figure 5, also listed below for discussion, are adopted from epidemiology and adapted for the purpose of assessing telemedicine.

- **Temporality/Gradient** simply implies that (1) the increase in access to medical care, however measured, follows the introduction/availability of telemedicine and is not the result of simultaneous reduction in travel costs through improvement in transportation or other reductions in opportunity costs; and, (2) the "stronger" the telemedicine program(reputation of consulting center, diffusion of information regarding program, acceptance by physicians/administrators) the greater will be the impact on access

{Note: the gradient criterion may imply a concomitant "temporal" gradient associated with the duration of the telemedicine program}

- **Statistical Significance/Strength**: A statistically significant correlation or association (excluding artifactual and indirect correlations) has been demonstrated between implementation of telemedicine and change in access. Particular emphasis should be directed toward those studies that provide statistically significant results. **Strength** is the degree of increase or decrease in access for persons using telemedicine versus those who use in-person care for the same conditions and the same level of illness severity. Alternate measures include various aspects of opportunity cost, including travel costs, travel time and distance, lost wages, and convenience.

- **Consistency** pertains to the degree of consensus on the impact of telemedicine on access obtained from a wide variety of places, settings, and applications as well as agreement of findings with currently accepted understanding of distance-interaction processes. Hence, the rules of evidence are summarized in terms of empirical findings, the sequence of variables, the strength of the association and the consistence with other studies. If an application is effective consistently across a representative set of indicators, it is not necessary to evaluate all indicators. (Lewin 2000) This same notion may apply to the study of access in that the weight of findings of increased access for a wide variety of applications in diverse geographical and institutional settings can provide reliable evidence regarding the effects of telemedicine on access.

- **Coherence** in the telemedicine-access equation means that the increase in accessibility is compatible with existing theory and knowledge. This presupposes a theoretical basis for the assumed relationship between telemedicine and increased access, which is derived from the ecology and ethology of human behavior.

The *Principle of Least Action*, sometimes called the *Maupertuis Principle*, after the French philosopher, mathematician and geodesist (1698-1759) is one of the greatest generalizations in all physical science. Although not fully appreciated until the development of quantum mechanics in the early 20th Century by Heisenberg and von Neumann,[1] (Stoltzner 2000) The "principle of least effort" forms the basis of general laws of physics, as well as assessments of German military strategy in WWII, behavior of

[1] Maupertuis annunciated the principle in 1746 in hopes that it might unify the laws of the universe and combined it with an attempted proof of the existence of God. In part the principle read: "The laws of movement thus deduced {from the Principle of Least Action}, being found to be precisely the same as those observed in nature, we can admire the application of it to all phenomena, in the movement of animals, in the vegetation of plants, in the revolution of the heavenly bodies; and the spectacle of the universe becomes so much the grander, so much more beautiful, so much more worthy of its Author…"

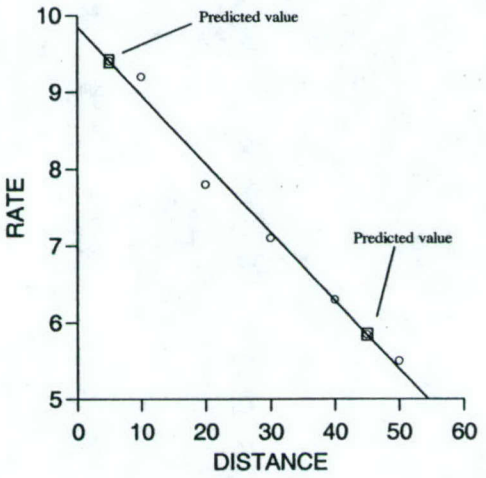
serial killers, addresses in cyberspace, and the use of words. (Harries 1999; Umstatter 2003).

Basically, the Principle states that "nature is lazy" and that nature (including humans) will satisfy its needs and desires with the least possible exertion. This translates directly to human travel behavior where it has been demonstrated, ceteris paribus, people use the opportunity nearest to them.

Plausibility- This notion is reflected in human behavior by the concept of the distance decay function (**Distance Decay Function Figure**). Simply put, the function describes a decreasing probability of interaction with a place with increasing distance from that place. The use of a distance decay formulation helps identify potential patterns of service utilization by people. Here too, there is considerable literature, especially in the mental health field where it is known as Jarvis' Law, that supports this proposition. Hence, we are on solid ground in terms of both the Principle of Least Action and the empirically defined law of distance-interaction function as a basis for implementing telemedicine.

Figure 6

Distance Decay Function



There is a considerable literature to support the notion that increased distance to medical care is associated with decreased utilization. (Guagliardo 2004; Gregory et al. 2000; Fortney, et al. 1999; Shannon et al. 1969) Further, other things equal, distance can be the determining factor as to when and if a person takes advantage of a medical care

opportunity almost regardless of the distance to medical care involved. (Gregory et al. 2000; Shannon et al. 1986) These empirical observations support the notion of plausibility and directly reflect the basic principles and laws pertaining to human movement behavior.

- **Specificity**: Ideally, the effect has only one cause. An outcome predicted by one primary factor adds credibility to a causal claim - a single cause is related to single effect. This would be attained when we are able to demonstrate that a single aspect of telemedicine is associated with a single effect on access.
- **Experiment**: The demonstration that under controlled conditions changing the exposure causes a change in the outcome – in this case – the conditions suggested earlier pertaining to the need for large scale, clinical trials.
- **Analogy**: Typically, we are more willing to accept arguments that resemble others we have already accepted. In other words, a commonly accepted phenomenon in one area can be applied to another area. With regard to telemedicine and access, this would again refer to the voluminous literature that supports the notion that placing a medical facility "nearer" a target population increases utilization by that population

In brief, we have proposed here a revised strategy for evaluating the effects of telemedicine based on the available evidence, with special emphasis on the association between telemedicine and access to care, as an illustration on the use of this strategy. The impetus for this effort derives from past, present and, most likely, future difficulties encountered in attempting to evaluate the multi-dimensional and continually evolving technology and applications of telemedicine. Moreover, it is important to note that telemedicine is not the only type of program that is difficult to evaluate using the controlled experiment. It may be recalled that the evaluation of the effectiveness of the Salk Polio Vaccine was based on time series analysis and not a controlled experiment. Suggestions pertaining to the need for looking toward triangulation as one means of not only resolving the problem but also providing more feasible, holistic, and "surer" analyses are found in other fields as well. We have argued here that triangulation may be a viable alternative evaluation strategy for telemedicine. Additionally, we have presented a preliminary and, perhaps, rather simplistic first attempt to illustrate the concept and how it might be implemented. To be sure, questions pertaining to the operationalization of the terms and development of appropriate measures remain to be answered. One central problem, should the strategy prove viable, is a mechanism to collect data from various sources, and a central repository for results from the various types of research methodologies, both quantitative and qualitative, that will form the basis for implementing this strategy.

Summary and Conclusion

Despite several decades of growth and deployment of telemedicine programs, a review of the literature illustrates that to date the majority of evaluation studies does not rise to the level of producing definitive results regarding the benefits and costs of this field. Hence, claims of telemedicine program efficacy pertaining to improved access, equal or enhanced quality when compared to traditional medical care; and, reduced costs, cannot be made with assurance. Reasons for this problem are many including:

- A failure to use a precise and uniform definition of "telemedicine" and the multi-dimensionality of the innovation bundle of telemedicine
- The continuous improvement of the underlying technology of telemedicine and the expansion of applications
- Experimental problems such as
    - Lack of clarity in specifying the experimental variables
    - Inability of experimental allocation and the ease of jumping experimental allocation
    - Lack of program maturity and steady state operation
    - Non-specificity of program affects, delayed effects, and unintended effects
    - Lack of large scale programs that would permit experimental studies
- Insufficient funding for large scale experimental studies
- Failure to exploit the full potential of telemedicine technologies

The inadequacies of scientific rigor aside, the overwhelming majority of published evaluations of telemedicine programs support the notion that telemedicine does improve access; is a valid substitute for in person care in many applications; and, does or has the potential to reduce opportunity costs. And, in spite of the lack of substantive evaluation research, telemedicine programs continue to proliferate but, perhaps, not to the extent that proponents feel is appropriate, given the presumed potential that should be demonstrable through evaluation. Proponents continue to call for support, while the field is criticized for not having adequate research evidence to justify its full adoption as an integral component of the health care armamentarium.

It appears that telemedicine program evaluation is faced with a quandary. Do we continue to pursue what appears to be a futile search for definitive results from small, limited studies that are scientifically rigorous? Or, do we devise a strategy or strategies for assessing all the available evidence in systematic, coherent and consistent fashion through triangulation?

In this paper, we have presented two strategies, which are not mutually exclusive. The first strategy is to fund large scale experimental telemedicine programs/projects that can be designed and implemented to collect data sufficient to test specific dimensions and effects of the technology. Data from such studies can then be assessed with statistical analyses to draw probable and credible conclusions. We proposed a three-dimensional model for evaluating telemedicine on the basis of the varied perspectives of the client, provider, and society at large, the specific clinical or other application, and the technological configuration. Moreover, as findings are reported, results can be accumulated, using this three-dimensional telemedicine data matrix, to permit the triangulation of results and the translation of results into a summative research format. In turn, the cumulative research findings would form the basis for reaching conclusions regarding the benefits and costs of telemedicine.

We also proposed and illustrated a second alternative using theoretical triangulation as one basis for assessing the impact of telemedicine on access to care. This strategy derives from the fact that in reality our search for the "holy grail" of scientifically evaluating telemedicine programs in the traditional manner must be revised. The new strategy is based upon established theory together with the cumulative data from research studies that may be based on imperfect designs. The criteria for judging the

results include: statistical significance, temporality, strength, consistency, coherence, plausibility, and experiment; as well as integrating results from both quantitative and qualitative research designs.

This forum, dedicated to the future of telemedicine, is an appropriate setting for these initial statements on and consideration of program evaluation strategies. We hope our suggestions and subsequent discussions about them eventually will lead to a resolution of the quandary now facing telemedicine program evaluation. We welcome your comments, criticisms, and thoughts on our proposals.

References:

Bashshur, R. L. (1980) "A Proposed Model for Evaluating Telemedicine." L. Parker and C. Olgren (eds.) Teleconferencing and Interactive Media University of Wisconsin-Extension Center for Interactive Programs 211-19

DeChant, H., Tohme, W., Mun, S., et al. (1996) "Health Systems Evaluation of Telemedicine: A Staged Approach." Telemedicine Journal 2:4 303-12

Drake, D. E. (2003) "Evaluating Telemedicine: A Literature Review." eHealth International

Fortney, J., Rost, K., Zhang, M., and Warren, J. (1999) "The Impact of Geographic Accessibility on the Intensity and Quality of Depression Treatment." Medical Care 37(():884-93

Fowler, T. (1998) "The Formality of Reality: Xavier Zubiri's Critique of Hume's Analysis of Causality." Xavier Zubiri Review 1:57-66

Guagliardo, M. (2004) "Spatial Accessibility of Primary Care: Concepts, Methods and Challenges." International Journal of Health Geographics 3:3
http://www.ij-healthgeographics.com/content/3/1/3

Gregory, P., Malka, E., Kostis, J., et al. (2000) "Impact of Geographic Proximity to Cardiac Revascularization Services and Service Utilization." Medical Care 38(1):45-57

Hailey, D., Roine, R., and Ohinmaa, A. (2002) "Systematic Review of Evidence for the Benefits of Telemedicine." Journal of Telemedicine and Telecare 8 (Suppl. 1): S1:1-7

Handfield, R., Ragatz, G., and Monczka, R. (1995) "A Triangulated Approach to Empirical Research: The Global Procurement and Supply Chain Benchmarking Initiative." Decision Line 26(3):

Harries, K. (1999) Mapping Crime: Principles and Practice Research Report Crime Mapping Research Center, U. S. Department of Justice

Health Canada (2000) Evaluating Telehealth 'Solutions' A Review and Synthesis of the Telehealth Evaluation Literature. Office of Health and the Information Highway http://www.hc-sc.gc.ca/-bsi/pubs/2000_tele/tele_e.html

Hersh, W., Helfand, M., Wallace, J., et al. (2001) "Clinical Outcomes Resulting from Telemedicine Interventions: A Systematic Review." BMC Medical Informatics and Decision Making 1:5 http://www.biomedcentral.com/1472-6947/1/5

Hill, B.A. (1965) "The Environment and Disease: Association or Causation?" Proceedings of the Royal Society of Medicine 58:295-300

Hilty, D., Marks, S., Urness, D., Yellowlees, P., and Nesbitt, T. (2004) "Clinical and Educational Telepsychiatry Applications: A Review." Canadian Journal of Psychiatry 49:1 12 -23

Jick, T. (1979) "Mixing Qualitative and Quantitative Methods: Triangulation in Action." Administrative Science Quarterly 74:602-11

Lewin (2000) Assessment of Approaches to Evaluating Telemedicine Final Report Prepared for Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services http://aspe.hhs.gov/health/reports/AAET/aaet.htm

Mackay, W. (2004) "Triangulation Within and Across HCI Disciplines." http://www-ihm.lri.fr/~mackay/pdffiles/HCI98.Commentary.pdf

Mackay, W. and Fayard, A. (1997) "HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines." In Proceedings of Designing Interactive Systems, DIS '97 ACM: Amsterdam, The Netherlands

Mair, F., Whitten, P. (2000) "Systematic Review of Studies of Patient Satisfaction with Telemedicine." British Medical Journal 329 June 1517-20

Milloy, S., and Gough, M. (2004) "The Environmental Cancer Epidemic That Never Was." Regulation: The CATO Review of Business and Government http://www.cato.org/pubs/regulation/reg19n2g.html

Mitchell, E. (1986) "Multiple Triangulation: A Methodology for Nursing Science." Advances in Nursing Science 8:18-26

Congress (2001) "2001 Report to Congress on Telemedicine: Evaluation and Research." Office for the Advancement of Telehealth Publications http://telehealtah.hrsa.gov/pubs/report2001/eval.htm

Roine, R., Ohinmaa, A., and Hailey, D. (2001) "Assessing Telemedicine: A Systematic Review of the Literature" Canadian Medical Association Journal 165(6):765-71

Rothman, K. (1996) "Causation and Causal Inference." Chapter 1 in D. Schottenfeld and J. Fraumeni (Editors) Cancer Epidemiology and Prevention Oxford University Press 2nd Edition

Shannon, G., Bashshur, R., and Metzner, C. (1969) "The Concept of Distance as a Factor in Accessibility and Utilization of Health Care." Medical Care Review 26:143-61

Shannon, G., Bashshur, R., and Lovett, J. (1986) "Distance and the Utilization of Mental Health Services." Milbank Memorial Fund Quarterly Health and Society 64(2):302-30

Shea, S., Starren, J., Weinstock, R., Knudson, P., et al. (2002) "Columbia University's Informatics for Diabetes Education and Telemedicine (IDEATel) Project: Rationale and Design." Journal of the American Medical Informatics Association 9:49-62

Stead, W. (2002) in Hersh, W., Patterson, P., and Kraemer, D. "Telehealth: The Need for Evaluation Redux." Journal of the American Medical Informatics Association 9:1 89-91

Stoltzner, M. (2000) "The Principle of Least Action as the Logical Empiricist's Shibboleth." Proceedings History and Philosophy of Modern Physics http://philsci-archive.pitt.edu/archive/00000616/00/shibboleth-archive.doc

Umstatter, W. (2003) "Das "Principle of Least Effort" in der Wissenshaft." http://www.ib.hu-berlin.de/~wumsta/infopub/lectures/leasteffort03a.pdf

Whitten, P., Mair, F., Haycox, A., et al. (2002) "Systematic Review of Cost Effectiveness of Telemedicine Interventions." British Medical Journal 324 June 1434-7

A Review of Clinical Outcomes in Telemedicine/Telehealth

Paul Heinzelmann
Christy Williams
Nancy Lugn
Joseph Kvedar

Partners Telemedicine
Boston, MA

A Review of Clinical Outcomes in Telemedicine/Telehealth

Purpose: To survey the field of published telemedicine literature and identify and report clinically relevant outcomes studies from 1996-present.

Objectives:

- Suggest and a conceptual framework for understanding telemedicine research within the larger clinical process
- Systematically collect, review, and describe the results of telemedicine studies that focus on clinically-relevant outcomes
- Discuss current evidence for the clinical benefit of telemedicine
- Identify study design issues
- Suggest recommendations for future research

Methods:

- **Phase 1.** Electronic literature review of peer-reviewed journals. Initial abstract search using PubMed and specified search topics for the period1996-present, and hand collection of other relevant articles. Our aim is to review, tabulate, and provide a descriptive and summative report.
- **Phase 2.** Incorporate feedback from symposium participants.
- **Phase 3**. Final manuscript compilation.

Outline (Table of Contents)

Rationale

Introduction

Geography no longer limits healthcare practitioners from seeing a radiograph, microscope slide, skin lesion, echocardiogram, or a patient's behavior. It no longer prevents a heartbeat or a patient's account of their illness from being heard; and it doesn't even prevent surgery for those in areas without a surgeon. Though not widely adopted, these examples demonstrate how the innovative use of information and communication technologies (ICTs) continues to evolve as a way of enabling access to quality and cost-effective healthcare delivery.

In an ongoing way, ICTs have increasingly, and often invisibly, integrated into the existing healthcare system. This integration has given rise to multiple terms aiming to describe this phenomenon. 'Telemedicine', 'telehealth', 'e-Health', 'telehealthcare' are terms that are expanding our vocabulary to describe this emerging field, and nearly every medical specialty and setting now has its own 'tele' counterpart that denotes where this evolution is taking place. While these new terms are often used to showcase innovative applications, they also remind us that these innovations remain separate from conventional and accepted methods of care. Until the time comes when we no longer need to signify the uniqueness of ICTs in healthcare delivery, we are tasked with demonstrating how this new way of doing things is as good or better than the status quo. As diffusion of this union between ICTs and healthcare continues to grow, so does the need for solid evidence of its benefit. It is likely that successful integration into our healthcare system will be marked not with fanfare but with the quiet extinction of the 'tele' and 'e' prefixes that we now so eagerly use.

What the prior reviews have shown to date

Hersh, et al, (2002) completed a systematic review of clinical outcomes up to February of 2001 focusing on two types of telemedicine applications: home-based and office/hospital-based. Of 58 published studies identified, 15 assessed clinical management / patient outcomes and 57 of the studies evaluated in the category of diagnosis. The studies identified were rated for quality and were categorized according to the degree of evidence supporting the telemedicine intervention. Although it was recognized that very few high-quality studies were identified, it was felt that the fields of psychiatry and dermatology had the strongest evidence for efficacy for both diagnosis and clinical management. Reasonable evidence that history and physical exam via telemedicine has good sensitivity and specificity was also noted. This review only examined those fields where the conventional intervention is an in-person exam; therefore the fields of pathology and radiology were excluded.

Hailey, et al., (2002) also completed a systematic review of the benefits of telemedicine as reported by studies between the years 1966 and 2000. A total of 66 articles were identified, of which, 46 assessed at least some clinical outcome. Fifty-six percent suggested that telemedicine had advantages over the conventional clinical intervention, 36% were inconclusive or had identified some negative aspect to the telemedicine intervention, and 8% of the studies found the conventional intervention had advantages over the telemedicine intervention. This review reported that the most convincing evidence existed for teleradiology, tele-mental health

(telepsychiatry), transmission of echocardiograms, teledermatology, home monitoring, and some medical consultations. Economic assessment and subjective rating about the study's potential to influence future decision-making on telemedicine services were also included.

Aoki, et al., (2003) evaluated outcomes and methods in telemedicine over a similar time period (1966-2000). Of the 112 twelve identified articles, 49 studies evaluated variables in the category of diagnosis, three within clinical management, and 26 reviewed patient satisfaction.

We will expand on prior reviews with a broader and updated examination of the literature.

Conceptual Model of the Clinical Process

To better understand telemedicine outcomes research, let us first consider the clinical process and how studies are typically designed to measure the impact of telemedicine as a clinical intervention. We will refer to the clinical intervention alternatives as: *conventional*, *telemedicine*, and *gold standard.*

Typically, a patient enters the clinical process with symptoms of some kind or because clinical risk factors for disease have been recognized and suggest the need for preventive screening. In this scheme, clinically-relevant information (history, physical exam, testing) is gathered by a clinician who then makes a diagnosis, or if uncertain, he/she consults with a *diagnostic specialist* (e.g. radiologist or pathologist*)* to arrive at a diagnosis. The *clinician* then formulates a prognosis and decides upon a management plan, or consults a *clinical specialist* (dermatologist, cardiologist, ophthalmologist, etc. *)* to assist with this. Management typically involves formulation of a treatment plan and may also include follow-up, monitoring, or a contingency component. At some future time point, patient outcomes can be measured which reflect the effectiveness of the clinical intervention. Finally, if widely adopted and sustained, the effects on a population can be observed, measured and compared.

In reality, the clinical process is not as simplistic as this description suggests. The clinical environment (e.g. available resources, financing scheme, policies, etc) and human factors (e.g. knowledge, attitudes, practices, beliefs) also play a significant role, and create great variability in care from setting to setting. In recent years, the push of "evidence-based medicine" has aimed to standardize the behaviors of clinicians to improve quality and clinical outcomes. Despite this, there remains great geographic variability in care on both a national and global scale and as a result, an important role for telemedicine emerges.

There are multiple variables that can be measured throughout this entire process and exactly what to measure isn't always obvious. Similarly, multiple theoretical frameworks have been proposed for the evaluation of telemedicine (Bashshur1998). We will be examining telemedicine outcomes from a clinical perspective, and for convenience, we will divide the clinical process into the following categories:

1. **Diagnosis**
2. **Clinical Management**
3. **Patient Outcomes**
4. **Participant satisfaction**

The clinical process and alternative clinical interventions are illustrated in Figure 1.

(insert Figure 1)

Identifying Key Clinical Research Questions

When examining the effect of telemedicine, it is common to compare medical specialties or care settings and note their "maturity" within the parallel field of telemedicine. While this approach does help identify where more research could be directed, it must be appreciated that each specialty or care setting has its own set of unique tools, functions and objectives within the wider process of clinical care. Table 1. illustrates how the various specialties can be further organized by clinical function, and in turn, suggest an expected focus for clinical research. This does not mean to imply that some categories of the clinical process (ie. diagnosis, clinical management, patient outcome, participant satisfaction) are to be disregarded for some specialties, but rather highlight those areas that are of particular importance to that specialty. (This table should be examined closely as we will refer to the described categories of the clinical process and the categories of clinical function determined by specialty, throughout this review.)

For example, it can be argued that, because the fields of radiology and pathology are primarily diagnostic in function, the most appropriate focus for outcomes research is in the area of diagnostic accuracy, as opposed to patient satisfaction. Meanwhile, specialties such as dermatology, however, have both a diagnostic and clinical management function, and as a result have a broader research agenda. Those applications that have a disease monitoring function must foremost demonstrate improved patient outcomes (e.g. decreased hospitalization, improved survival), and patient satisfaction.

Within the context outlined above, we will explore the existing evidence for telemedicine by specialty through a review of the literature and provide more specific information on what has been done and what needs to be done to encourage continued growth and maturity of telemedicine applications.

Literature Review Methods

This review employs a three-phase approach to gather evidence for the clinical impact of telemedicine. Phase 1 involves beginning a broad review of the field of telemedicine outcomes research and a description of our findings. Phase 2 involves the incorporation of symposium feedback. Phase 3 consists of final manuscript development.

Phase 1. Broad literature search from the years 1996-present.

Our aim is to provide a descriptive account of the results rather than attempting to systematically aggregate results with a meta-analysis approach. We arranged our search into three areas: 1) diagnosis, 2) clinical management and patient outcomes, 3) participant satisfaction. (Note: the categories of clinical management and patient outcome were combined for convenience).

Search strategy: Several key terms were searched using the PubMed web portal to identify published studies in peer-reviewed journals that focused on the clinical processes we have outlined in our conceptual model. It has been noted that the vast majority of telemedicine outcomes research has been completed after 1995 (Aoki, 2003), and that technology advancement may make past studies less relevant to present realities. We have therefore, limited our search to studies published from1996 to present.

PubMed provides access to bibliographic information that includes MEDLINE and some additional life science journals that receive a qualitative review by the National Library of Medicine, and contains bibliographic citations and author abstracts from more than 4,600 biomedical journals published in the United States and 70 other countries.

Literature Search Strategies were organized by categories of the clinical process.

Studies of Diagnosis

Search terms used for this section included 'telemedicine AND diagnostic accuracy' and 'telehealth AND diagnostic accuracy'. One hundred and sixty-six articles were identified, and the abstracts were reviewed to identify comparative studies that sought to measure a diagnostic outcome and compare it to a control of some kind.

Inclusion criteria: We included only studies that compared telemedicine to a control intervention of some kind.

Exclusion criteria: Case studies, feasibility studies, narrative reviews, and anecdotal reports were removed from the analysis, as were preliminary reports of later published studies.

Of the 166 abstracts, 116 (70%) met the criteria and were tabulated for analysis with simple descriptive statistics. Abstracts and articles for review were also identified by hand searching literature cited in the references section of several review articles bring the total to 150 studies.

Studies of Clinical Management and Patient Outcome

One-hundred twenty one abstracts were identified with the search terms 'telemedicine AND medical outcome, 'telehealth AND medical outcome', and 'telehealth AND clinical outcome'. Additional articles for review were also identified by hand searching literature cited in the references section of several review articles. Ultimately, a total of 33 abstracts met our criteria of being comparative to a control intervention of some kind.

Inclusion criteria: We included only studies that compared outcomes of a telemedicine intervention to a control intervention of some kind.

Exclusion criteria: Case studies, narrative reviews, and anecdotal reports were removed from the analysis as were preliminary reports of later published studies. We also excluded studies that did not compare the telemedicine intervention to an in-person control intervention (e.g. radiology and pathology).

Studies of Participant Satisfaction

PubMed MEDLINE was used to search the key words 'telemedicine and patient satisfaction', 'telemedicine AND quality of life', and 'telemedicine and provider satisfaction'.

Using only the abstracts, setting was determined and grouped into the following categories: acute care, long-term care, home monitoring, home health, and other. Acute care included hospitals, emergency departments, and clinics. Long-term care included rehabilitation hospitals and nursing homes. The 'other' category included education and all studies evaluating the technology. Studies were also grouped into discipline and the number of participants. If a study had 2 disciplines, they were counted in each separately. For studies with 3 or more disciplines, they were considered 'multi-discipline.' If a single study resulted in multiple publications, only the most recently published was included in our evaluation. Once data was compiled, four articles were hand-searched to find studies cited in the bibliographies that were not identified with the original search. A total of 168 studies were identified.

Results

Following are the results we achieved using the protocols outlined above. These results expand on those of other systematic reviews of the literature (Hersh, Hailey, Aoki). Results are reported below by category of the clinical process as follows: 1) diagnosis, 2) clinical management and patient outcomes, and 3) participant satisfaction.

The use of small sample sizes and the reporting of data in a variety of formats makes aggregation of the results difficult, however, we can make some general observations by specialty. Table 2 provides an overview of the number and sources of the studies identified and reviewed to date.

(Insert Table 2)

Studies of diagnosis

Twenty-one specialties were identified among the 150 abstracts identified in our search. Table 3. provides a summary of the studies within the category of diagnosis, and a detailed reporting of the results can be found in Tables 5-27.

(insert Table 3)

Figure 2 illustrates the relative number of diagnostic studies by specialty (Only one study of diagnostic accuracy was identified for the fields of neonatology, neurosurgery, ambulatory care, and microbiology and these are not included in the figure.).

(insert Figure 2)

Of the studies that compared the telemedicine intervention to 'in-person exam', most were in the field of dermatology (27%). Of these dermatology studies, 65% were store-and-forward consultations while the remainder involved interactive consultations.

Not all the studies were supportive of the telemedicine intervention studied. To further estimate the impact of telemedicine interventions, each of the studies was reviewed and categorized by the authors as either 'supportive', 'inconclusive' or 'unsupportive' of the telemedicine intervention being studied. Supportive studies were those that had results and conclusions that were felt to indicate that the diagnostic capability of the telemedicine intervention was as good or better than the control it was being compared to; or there was improved diagnostic capability by enabling the opinion of a diagnostic specialist with a higher level of expertise. Inconclusive studies were those that were judged to be weakly unsupportive and of weak design, and unsupportive studies were those that indicated that the diagnostic accuracy was clearly inferior to the control intervention.

Of the 150 diagnostic studies evaluated, 79% were rated as 'supportive' of telemedicine, 15% were rated as 'inconclusive', and 7% were rated as 'unsupportive'. Interestingly, 62.5% of the 'inconclusive' or 'unsupportive' results involved interactive applications, suggesting that further research into the diagnostic accuracy of interactive applications may be warranted, particularly for pathology and dermatology. The distribution of inconclusive and unsupportive studies can be seen in Table 4.

(insert Table 4)

At this time, we can examine several of the specialties in more depth. As described initially, these are again categorized by clinical function and within these headings, by specialty..

Specialties with a Diagnostic Function

Radiology:

This is arguably the field that has most successfully integrated telemedicine interventions into everyday practice. Digitized radiographs, CTs, MRIs are now commonly available through PACS systems and are the standard in many hospitals. The number of radiologic studies may be surprisingly sparse given its unique status. However, this is likely due to the fact that teleradiology has already been widely adopted and the significant overlap with other specialties that rely heavily on radiologic diagnosis as part of their clinical function (eg. Emergency care, Neurology), and because of the clear cost-effectiveness, which mitigates the need for further study.

Nonetheless, noteworthy observations have been made through this review. One study demonstrated that radiograph images created by flatbed scanning provided equivalent diagnostic accuracy when compared to film radiographs, and superior accuracy when compared to images obtained with a digital camera. Other studies, however, showed that static radiographic images from a digital camera can be used effectively, but depend on having high quality original films, appropriate views, and zoomed views of the areas of interest. Image compression ratios of 20:1 do not appear to compromise diagnostic value. Also, increased level of expertise improved diagnosis for those with less experience or those in a different specialty/sub-specialty. As in

dermatology and pathology, previously determined "difficult" diagnoses resulted in a marked decrease in diagnostic accuracy and sensitivity when viewed in digitized form.

Pathology:
Our review revealed more diagnostic studies in the field of pathology than any other specialty. This reflects the 'natural fit' of telemedicine to the transmission of light microscope images. Most studies have compared diagnostic accuracy of telemedicine to examine frozen sections, intra-operative specimens, paraffin sections and other fixed samples. A smaller number of studies evaluated the viewing of fine needle aspirations (FNA) (e.g. breast, prostate, lung). Evidence for screening of cervical smears is conflicting, but generally supportive. Studies also conflict on whether dynamic-robotic telepathology can adequately provide an alternative to on-site pathologists. As with dermatology, current evidence seems to indicate that live video applications need further research to demonstrate diagnostic accuracy. Previously identified "difficult" specimens resulted in notably less diagnostic accuracy by telemedicine.

Specialties with a Diagnostic and Clinical Care Function

Dermatology:
Of all the fields that have both a diagnostic and clinical management function- dermatology has completed the most studies of diagnostic accuracy. Overall, agreement between dermatologists ranged from 51-100%, and in general improved when medical history was included. As expected, diagnostic accuracy of skin lesions depends on the reported confidence of the clinician and the degree of "diagnostic difficulty". Image quality rating showed at least a modest effect on diagnostic accuracy. Diagnostic accuracy appears to be lower for identification of benign skin tumors and some rashes, and generally higher for pigmented lesions and the detection of skin cancer. It has also been shown that diagnoses made by GPs are improved by consulting dermatologists using telemedicine. Like pathology, store-and-forward applications have more supportive evidence for benefit than interactive applications.

Cardiology:
Most studies evaluated the diagnostic accuracy of transmitted echocardiograms with interactive telemedicine, and the majority demonstrated high concordance particularly in the ability to exclude major congenital heart disease in neonates. There is some evidence that it may provide a means for rapid screening of after-hours patients with chest pain through consultation of distant cardiologists, but this requires further study.

The ability to distinguish functional versus organic murmurs in pediatric patients with transmitted heart sounds from a electronic stethoscope showed mixed results, but appears to be most accurate with children over 5 years of age and with experienced clinicians.

Ophthalmology:
Half of the diagnostic studies in ophthalmology were judged to be 'inconclusive', and one of which was 'unsupportive'. The majority of studies evaluated fundoscopic exam, which was shown to be limited in diabetic and HIV positive patients with cataracts. Diagnosis by slit lamp was examined in two studies and results varied broadly by diagnosis. In general, more research is needed before conclusive results can be drawn.

Psychiatry:
All studies involved live interactive applications and were supportive of the telemedicine alterative to in-person consultation. Specifically, neuropsychiatric testing, diagnosis of dementia, and to a lesser degree, use of the mini-mental status exam (MMSE) were effectively completed through telemedicine. These studies were limited however, by their relatively small ample size.

Emergency care:
Of the 6 studies identified in this review, 5 involved live-interactive telemedicine. In general, diagnostic accuracy within these few studies ranged from 70-99% and improved diagnosis sensitivity when used for consultation. (It should be noted that other studies that examined emergent care are also included in other sections such as radiology, urology, cardiology, etc.)

Specialties with a Disease Monitoring Function

Home-based care:
Five studies evaluated diagnostic accuracy in home monitoring situations. Blood pressure measurement accuracy is of particular value because ambulatory or home measurements are considered better predictors of cardiac risk than clinical measurements. The largest study in our review showed that home BP measurement with telemedicine more closely reproduced ambulatory BP measurement than those obtained in a clinical setting. Vital sign reproducibility and spirometry were also examined in two small studies, which demonstrated the accuracy of unsupervised home measurements of heart rate and spirometry.

Studies from other specialties were reviewed, but because fewer than five studies exist for each of these, they will not be discussed here. The results can be viewed in Tables in the Appendix however.

Studies of Clinical Management and Patient Outcomes

We have identified more studies than previously published literature reviews. It should be noted that studies examining outcome variables within this category are few in number, and when they do exist they are often examined secondarily to other outcomes such as diagnosis or participant satisfaction, and tend to focus on intermediate or surrogate measures. As a result, the results we have collected likely represent a somewhat limited sample.

Through our literature search protocol described above, we have identified 31 articles that examine a clinical outcome in comparison to some kind of control group. The majority (76%) of these studies involved an interactive mode (i.e real-time or near real-time with readily available communications between participants) and most have been in the area of home telehealth, followed by dermatology. See Table 28 for an overview of these studies, and Table 29 for a detailed account.

(insert Table 28, Table 29)

While none of these studies were considered 'unsupportive' of telemedicine, 5 were rated as 'inconclusive' indicating that the strength of evidence for the clinical benefit of telemedicine was weak.

(insert Table 16)

Noteworthy results for those specialties that have a clinical function of 1) diagnosis and clinical management and 2) disease monitoring are reported below. (Specialties with a clinical function of diagnosis (radiology and pathology) were not reported.

Specialties with a Diagnostic and Clinical Care Function

Dermatology:
Two studies evaluated interactive video to in-person exam and found that the same management plan was made in 64% of patients and the recommendation for biopsy was in agreement 86% of the time. One study of store-and-forward teledermatology demonstrated 90% agreement in management when images and history were provided in one study, while another showed a range of 56-77% agreement between in-person and teleconsultants. The inter-observer treatment agreement between in-person examiners was 77% and ranged from 64 to 83% among teleconsultants. Another study comparing dermatologist to general practitioner showed that only 31% of dermatology cases could have been appropriately managed by a general practitioner, and demonstartes the potential value added by teledermatology. (Loane, Phillips, Zelickson, Whited, Taylor).

Multi-specialty Consultations:
Several studies evaluated programs that provide multiple teleconsultation services. In particular, one ambitiously attempted a multicenter, randomized controlled study involving over 2000 patients; however, published data describes only the protocol and patient demographics for the study rather than reporting actual clinical outcomes. (Wallace). Another reports decreased CHF hospitalizations and improved glucose control in diabetics. One outpatient store-and-forward study reported that 79% of store-and-forward cases had additional treatment recommendations, and 93% of the in-person cases had additional treatment recommendations which reflected a broad spectrum of recommendations among the two groups. Finally, online multi-specialty consultations (90% cancer care) resulted in 90% change in treatment recommendations. (Kvedar)

Emergency Care:
One study evaluated the benefit of teleradiology and found that it changed the initial treatment plan in 26% of cases. Another showed it could decrease the need for transport of prisoners to the ED in 64% of cases and decreased the time of the consult from 2.75 hours to .5 hours (Lee, Ellis).

Psychiatry:
Management agreement was 96% in one study, while another demonstrated no significant difference in global assessment functioning (GAF), and that telepsychiatry visits had higher attendance and were shorter in length. (Zaylor, Elford)

Cardiology:
One study compared the recommended rate of follow-up echocardiogram for telestethoscope and traditional in-person care and found that the sensitivity and specificity was 88% and 97% respectively. The use of faxed ECGs to a cardiologist in decision-making about thrombolytics when compared to in-person exam with a resident physician was 87.5%. (Belmont, Srikanthan)

Ophthalmology:
A small study of a store-and-forward application revealed 42% agreement in management plans. (Schwartz)

Wound care:
Photographed images were compared to in-person exam in 24 patients. The sensitivity and specificity of management decisions was 87% and 78% respectively (Wirthlin)

Urology:
A study of interactive teleconsultation for possible urolithiasis in real and simulated patients showed that treatment plans changed in 50% of real patients and 17% of simulated patients (Hays).

NICU:
Tele-echocardiogram was compared to the convention of mailing a videotape of the ECHO in one study and demonstrated a reduction in NICU length of stay by 5.4 days, but was not significant in the sample used (Rendina).

Dentistry:
Store-and-forward oral imaging compared to in-person exam showed a kappa score of .93-1.0 for agreement for extraction and filing. (Patterson)

Specialties with a Disease Monitoring Function

Home-based care:
Perhaps the strongest evidence for a beneficial impact of telemedicine has been in the area of home-based monitoring (i.e. home telehealth). Over one-third of studies identified in this review involved some application to home-based care. There were several areas of focus among these studies.

Chronic Disease: Three studies examined home-based care programs that were not specific to one disease, but to multiple chronic and complex conditions that included CHF, diabetes, and COPD. These studies reported benefits such as improved perception of functional status, decreased ER visits, hospital stays, and quality of life. One study noted no difference in certain quality indicators (e.g. medication compliance, patient knowledge of disease, ability for self-care), satisfaction, or service utilization. Hospital admissions were decreased by 63% in a VA home monitoring program. (Ryan, Cherry, Johnston, Myer).

Diabetes Management: Two studies effectively evaluated the effect on diabetes management and reported significantly lower HgbA1C levels than controls, fewer hospital admissions, fewer ER visits, and fewer outpatient visits. Due to sample size, only decreased outpatient visits were statistically significantly (Bellazi, Cherry).

Heart Failure Monitoring: One study demonstrated statistically significant outcomes were identified at 3 months in one program and included fewer re-admission rates, shorter hospital stays, and decreased charges (Benatar). Another study found no group differences for functional status, depression, or health-related quality of life (LaFramboise).

Blood Pressure Monitoring: The demonstrated value of ambulatory and home blood pressure monitoring over office-based measurements (see diagnosis section) make this application particularly important. One study that employed a computer-controlled telephone reminder system significantly improved medication adherence and diastolic BP control. (Friedman).

Pulmonary/Asthma management: A small study in children demonstrated improved inhaler technique scores, peak flow values, and quality of life as reported by caregivers. No significant differences were found in clinical outcomes for lung transplant candidates using home spirometry. (Mullen, Bruderman).

Studies of Participant Satisfaction

A total of 168 studies were found and abstracts reviewed, and the results will be reported broadly in this category, rather than by clinical function. The majority of the settings were acute-care (64.1%), followed by 'other' (16.2%), home monitoring (12.0%), home health (6.0%), and long-term care (2.4%). A compilation of 25 fields with only 1-9 papers published constituted the majority (48.9%). Psychiatry, with 32 articles, is the most study field in regards to patient satisfaction (18.4%). Twenty-eight articles were multi-disciplinary (15.5%). Dermatology followed with 13 articles (7.5%). Approximately two-thirds of studies had fewer than 100 subjects (69.0%). Within those studies, 35.4% had fewer than 20. Though 24.1% overall had fewer than 20 subjects, 31.7% had more than 100. A large proportion of the studies used a survey or questionnaire (many using a Likert-Scale) to rate patient satisfaction, rather than anecdotal methods. Overall, patient satisfaction was high. Telemedicine was rated > 4.0 on a 5.0 scale in all studies given the results in the abstract. Satisfaction reported as a percentage of participants ranged from 55% to 100%. Between 20-40% (73%) of subjects preferred face-to-face versus telemedicine. One study with 585 subjects found that 70% preferred telemedicine to traveling for conventional care. In most studies where travel was a factor, patients had a higher preference for the telemedicine modality. One hundred eight were live-interactive modality, 11 were store-and-forward, and 3 used both modalities.

In a small study by Bratton et al, 94% of patients did not believe the technology had a negative effective on their relationship with the health-care provider; Callahan et al, in the field of psychiatry, found no significant differences between LI telemedicine and FTF in regards to a 'self-reported ability to speak freely.

The research suggests that travel is a significant burden for patients, and most would choose convenience when all other factors are equal. The psychiatry study by Callahan suggests that telemedicine has no effect on the interaction between providers and clients. Another psychiatry study by Chae et al assessed patient acceptance based on comfort, ease of self-expression, and quality of interpersonal relationship experienced with live-interactive telemedicine. All were found to be present at high levels. Cheung et al, studied live-interactive cardiology consultation, and found that patients stated they could communicate effectively with the cardiologist. In another study 17/41 patients felt more difficulty in speaking to the provider (Jones et al)

In general, providers appear to be less satisfied than patients, however, a generalized acceptability exists among patients, particularly in settings where it will most likely be implemented (remote/rural areas); current research establishes patients' willingness and satisfaction with live interactive telemedicine. However, this is not surprising given the similarity between live interactive and conventional care. With only 11 of the studies being store-and-forward, this opinion cannot be extrapolated to this area of telemedicine. However, with store-and-forward the care remains with the primary care physician (PCP), so provider satisfaction is more important and needs more research.

Many studies comment on the willingness of patients to use telemedicine again, a high predictor of acceptability.

Discussion

There appears to be very limited clinical evidence for the benefit of telemedicine interventions on clinical outcomes. To date, most studies have focused on diagnosis and patient satisfaction and relatively few examining clinical management and patient outcomes. As a result, certain specialties have achieved a greater level of maturity within telemedicine. In general, for those specialties that encompass diagnostic and clinical care functions (e.g. dermatology, emergency care, sub-specialties, etc) strong evidence for efficacy and effectiveness are needed within the areas of clinical management and patient outcomes. Significant challenges remain and higher quality studies are needed.

In general, it is agreed that stronger evidence is needed to demonstrate the benefits of telemedicine, and that studies that compare a telemedicine intervention to a conventional or gold standard intervention is required. Outcomes measured with a randomized controlled trial (RCT) are considered the strongest form of evidence. In a systematic review by Roine (2001), however, only six RCTs were identified within telemedicine clinical research literature, and these varied widely in quality. This review used the classification of study design by Jovell and Navarro-Rubio to stratify study design by strength of evidence. This classification is outlined in Table 30.

(insert Table 30)

Obstacles to conducting rigorous research in telemedicine

Numerous authors have noted design weaknesses in telemedicine evaluation research (Bashshur 1998), and have made recommendations for improving these research methodologies

(Yellowlees 1998, Holle 1999). If we explore this further, it appears that there are inherent operational characteristics that create obstacles to the completion of rigorous evaluation studies.

According to Holle (1999) these include:

1. Few large-scale, well-established programs in existence
2. Operations confined to remote, sparse populations
3. Rapid advancement of technology
4. Lack of technical standardization
5. Limited adoption and diffusion of telemedicine innovations into clinical routines

Hailey (2002) notes that these factors along with other circumstantial limitations result in studies with the following deficiencies:

- Small sample sizes
- Limited physician and patient study enrollment
- Limited ability for multi-center collaboration
- Results that are no longer relevant
- Use of retrospective data
- Different characteristics in study and control populations
- Poor response rates to surveys & high drop out rates
- Weak surrogate measures of efficacy/effectiveness, potentially confounded by other factors
- No follow-up on health status

In general, actions that can affect those circumstantial or operational factors, will improve the ability to conduct studies with appropriate statistical power to demonstrate an effect on the clinical process. To a large degree, these are dependant on the adoption and diffusion, and can be examined at the economic, individual, institutional, and societal levels. (Grigsby, et al 2002).

As Hailey (2002) observes, however, in addition to design factors that are imposed by the circumstances noted above, there are also study design and reporting failures introduced by investigators themselves.

Those introduced by investigators:

- Inadequate details of study design
- Limited details about control groups
- Details of some components of cost not given
- Incomplete data given on patient management
- Cost and/or effectiveness estimates based on the most optimistic scenario or on a subset of available data
- Narrow perspective in discussion on outcome measures
- Not taking into account the features of the health service that telemedicine is supporting
- Statements in the discussion and conclusions that are not supported by data and analysis given earlier

Unfortunately, there is no single optimum design, and methodological rigor needs to be balanced with the practicalities of resources and data availability.

Considerations and recommendations

Studies of Diagnosis

The reporting of outcomes in the category of diagnosis can take several forms. *Diagnostic accuracy* or *diagnostic agreement* typically refers to a comparison made between the diagnoses made through a telemedicine intervention and that made through the conventional clinical intervention or the 'gold standard'. This is usually reported on a nominal scale as a percentage that can be calculated from all paired ratings as *concordance* or *overall agreement. (Effective agreement* is occasionally used in such studies to minimize the role of chance agreement that would occur with rare diseases. This is calculated only on those paired ratings where at least one clinician detects an abnormality.) To more fully correct for chance, a *kappa coefficient* can be calculated, where a score of 1 indicates perfect agreement. This provides a measure of agreement across categories in a scale such as "mild', 'moderate', or 'severe'. *Inter-observer reliability* can also be reported and reflects consistency among diagnosticians. This latter measure is usually analyzed with ROC analysis. *Sensitivity* and *specificity* are also often reported in studies evaluating diagnostic capability.

Diagnostic accuracy is arguably not a direct measurement of a clinical outcome and as such, it may more accurately be considered a 'process indicator' or 'predictor' of an actual clinical outcome, as it is generally assumed that accurate diagnoses result in effective and appropriate clinical management, which in turn, result in the most beneficial outcomes.

Several study design issues are considered here.

Study Objectives

The objective of the study must be clear and should dictate the particpants in the study. The study should distinguish between showing *improvement in diagnosis* versus demonstrating that there is *no deterioration in diagnosis* due to the telemedicine intervention. Consider the following examples.

> *Example 1. Clinicians with differing specialty or experience*
> A study examining diagnoses of digitized head CTs may report *diagnostic improvement* by comparing the diagnoses of a remote emergency physician viewing a film radiograph in the conventional way with those made by a radiologist viewing the digitized version of the CT scan at a distant location. To be more complete, each of these could then be further assessed for *diagnostic accuracy* by comparing them to a gold standard diagnoses (i.e. made by a consensus group of neuro-radiologists). *Inter-observer agreement* of those in the consensus group or *reliability* could also be reported if desired.
>
> *Example 2. Clinicians within a single specialty*
> A study of the diagnoses made by a dermatologist viewing a digital image of a pigmented lesion versus the diagnoses made by a dermatologist through an in-person exam can measure *diagnostic agreement.* For more completeness each

could then compared to a gold standard diagnosis made by a biopsy and histological report.

Holle (1999) also makes the following observations about study design.

1. Participant Selection

Inclusion and exclusion criteria should be established prior to the study and avoidance of convenience samples or "suitable" patients is recommended. Excluding "difficult diagnoses" can create an obvious selection bias, which overestimates diagnostic accuracy..

2. Use of Diagnostic Categories

Diagnostic categories that are clinically-relevant should be used when possible. Including diagnostician confidence, or certainty allows the results to be stratified in a meaningful way, and prevents the exclusion of cases due to poor-quality, which can also lead to overestimation of diagnostic accuracy.

Intermediate variables that can be measured in addition to diagnostic accuracy or improvement in diagnosis include:

Image quality

- Resolution
- Compression
- Color characteristics

Diagnostician-related variables:

- Experience
- Diagnostic Confidence
- Specialty
- Degree of interactivity/control

Image "appropriateness"

- Source of image
- Inclusion of area of interest

Others

- Time to diagnosis
- Rate of deferred diagnosis
- Reported differential diagnoses
- Ability to identify specific pathologic findings

3. Gold Standard Intervention

Diagnoses made by a gold standard should be judged independently of the other diagnoses (e.g. blinded if possible). Note that gold standards that are defined as "expert opinion" are subject to inter-observer bias. Finally, in cases where only a subset of the study is subjected to the gold standard, a "work-up" bias is introduced. This can be adjusted for by subjecting a random sample of negative cases to the gold standard as well.

4. Sample Size

Diagnostic sensitivity will also not be accurate if there is a low prevalence of disease in the study sample. As a result, there is often a trade-off between statistical power and using a naturalistic clinical setting.

Further recommendations:

5. Addressing Confounding Variables

Generally, it must be recognized that there are many potentially *confounding and extraneous variables* that cannot easily be controlled for within and between the various aspects of the diagnostic process. Strategies such as matching, stratification and randomization can be used to minimize the effects of confounding variables.

6. Control groups

Studies designed to examine variables in such complex systems are often considered "quasi-experimental" in that they lack the full control of a true experimental design. Many diagnostic accuracy studies are quasi-experimental and as such, they often do not have a separate control group, or if there is one, subjects are not randomly assigned to it. Accordingly, many diagnostic accuracy studies utilize an *interrupted time-series design*, which can be defined in the following way:

> *A single group of subjects is subjected sequentially to two or more interventions (or to a single intervention and a no-intervention phase), with intervening measurement of appropriate outcome variables. There is no separate control group (and therefore no randomization) – the subjects act as their own control. The intervention variable is manipulated. (Sim, 2000)*

Studies of Clinical Management and Patient Outcomes

The studies included in this review are summarized in Table 17. Overall, conclusive results here are limited by the few number of studies conducted, study designs employed and the small number of subjects in many of the studies. With these limitations, it appears that at this time home-based monitoring (home telehealth) has the most evidence for beneficial impact on clinical management, particularly for general chronic disease management, CHF monitoring, and diabetes control. Dermatology, psychiatry, emergency care, cardiology, neonatal tele-echocardiography, wound care and dentistry appear to have promising results with some applications but additional studies are required. Unfortunately, one large-scale study, which employed over 2000 subjects in a multi-centered, randomized controlled trial has yet to report outcomes in the published literature. (Wallace).

An important focus of comparative outcomes studies is to demonstrate an impact on patients who were subjected to the telemedicine intervention, became healthier or achieved a certain degree of health status. The "effect" or outcome is usually a quantitative improvement of some

measurable criteria. There are multiple variables that may be measured as indicators of clinical outcome, but to date the most frequently reported outcomes are intermediate or surrogate variables in the clinical process.

Again, Holle (1999) notes that demonstrating a *causal relationship* presupposes internal validity, where the effect noted can't be attributed to confounding or extraneous random influences. To demonstrate this, however, a highly controlled study setting is required. In such a setting *clinical efficacy* may be reported, but it must be recognized that in order to demonstrate *clinical effectiveness,* telemedicine interventions must be tested in routine clinical settings where the results can be generalized. In summary, there is generally a trade-off between internal and external validity.

Telemedicine efficacy trials have inherent limitations (e.g. unable to use double-blinding or placebo), but the following are methodological issues for assuring internal validity (Holle 1999):

1. Choice of the Control Group
Pragmatic trials comparing the "best available standard" should be utilized.

2. Randomization
Though randomization is impractical in many situations, it is considered the method of achieving the strongest evidence for efficacy. Because creating a parallel study group not using a telemedicine intervention in a setting at the same time has logistic obstacles, an alternative to randomization of patients could be randomization to institutions that have implemented telemedicine.

3. Comparability of Nonrandomized Groups
Most trials evaluating telemedicine use nonrandomized designs that have historical or concurrent control groups. Strict inclusion criteria for both the control and intervention group should be adhered to.

4. Standardization of the Intervention
Including multiple institutions in a study can increase sample sizes, but it is important to assure a standardized intervention as much as possible.

5. Outcome Criteria
These may be objective or subjective.

There are many possible variables to measure throughout the category of clinical management and patient outcome. Assessment of the following end points of clinical care remain largely unexplored for the most part in the field of telemedicine.

- mortality, morbidity
- survival rates
- quality-adjusted life years (QALYs), disability adjusted life years (DALYs)
- health-related quality of life
- functional status

- disease-specific indicators of quality care

Instead, most studies have focused on intermediate variables, which may also be considered surrogate or 'process indicators' of clinical outcome. These include the following:

- change in medication
- change in treatment recommendations
- change in prognosis
- follow-up rates
- adherence to evidence-based protocols
- referral rates
- transfer rates
- time of visit
- provider-provider interactions
- patient-provider interactions
- patient satisfaction
- provider satisfaction

Monitoring or follow-up is an important part of clinical care that may be considered separately. Typical outcomes include:

- hospital admission rates
- ER visits
- other resource utilization indicators

Studies of Participant Satisfaction

The research suggests that travel is a significant burden for patients, and most would choose convenience when all other factors are equal. Many studies also comment on the willingness of patients to use telemedicine again, a high predictor of acceptability.

A prior review states that the provider-patient interaction hasn't been studied thoroughly. However, Callahan, in the field of psychiatry, suggested that telemedicine had no effect on the interaction between providers and clients. Another psychiatry study by Chae et al assessed patient acceptance based on comfort, ease of self-expression, and quality of interpersonal relationship. They also found the interaction to be good. Both studies used live-interactive telemedicine. Cheung et al, based on a live-interactive cardiology consultation, found that pts stated they could communicate effectively with the cardiologist. In another study 17 of 41 patients felt more difficulty in speaking to the provider (Jones et al). Also in support of a physician and patients' ability to establish rapport is the multitude of positive satisfaction studies supporting telemedicine in the field of psychiatry, an area requiring relationships and interpersonal communication skills to be successful.

Such a generalized acceptability exists among patients, particularly in settings where it'll most likely be implemented (remote/rural areas), current research establishes patients' willingness and satisfaction with live interactive telemedicine. With only 11 of the studies examining store-and-forward services, participant satisfaction remains uncertain. Due to the fact that store-and-forward applications tend to be provider-to-provider transactions, provider satisfaction is of particular relevance here. Among the few articles that examined provider perceptions, they generally seem less satisfied than patients, and undoubtedly likely plays a strong role in the adoption of the technology.

Questionnaire design is of particular importance in assessing satisfaction. If one is not using a validated and standardized questionnaire, the following recommendations should be considered. (Adapted from Holle 1999).

1 Use clear wording
2 One question per phrase
3 Use scales with four labeled categories for asymmetric questions, and scales with five, seven, or visual analog scale (VAS) for symmetric questions.
4 Use consistent response formats
5 Always conduct a pretest

The challenge of integrating evaluations into service delivery

The Institute of Medicine's 1996 report on telemedicine evaluation recognized the value of rigorous and systematic evaluation, and that evidence gained by this approach is important for adoption (Field, 2002). The highly structured approach typified by clinical trials and evaluation frameworks such as health technology assessment (HTA) mandate the production of evidence for telemedicine's benefit (Williams 2002). Nonetheless, it has been recognized that certain conflicts may arise when demanding evaluation protocols are introduced into the socially and politically influenced clinical practice environment, particularly early in a technology's implementation or when technologies are fast-changing.

Hersh et. al., (2002) advocates for randomized controlled trials, and states that studies should seek to provide adequate statistical power. Due to the challenges of integrating evaluation with practice, however, he also advocates for the employment of newer systematic research methodologies such as "tracker trials" (see Lilford 2000) to accommodate new and/or rapidly changing interventions and compare efficacy.

Finch, et al. (2003), also recognized that combining evaluation with telemedicine service development has unique challenges, and that the structure necessary for rigorous evaluation is often difficult to introduce into a clinical setting and may be disruptive to the clinical process that is being evaluated. For instance, adding telemedicine alternatives into the clinical workflow may increase the perceived need for risk management among clinicians because the safety and effectiveness of these new alternatives remain unsubstantiated. More pragmatic approaches that use qualitative and quantitative methods are required to improve the quality of such research.

This issue has already been observed with the integration of patient care information systems into clinical settings. Berg (1999) advocates an iterative or socio-technical approach "in which the distinctions between 'analysis', 'design', 'implementation', and 'evaluation' blur." User-oriented perspectives that emphasize the realities of the workplace should be the starting point for design and implementation.

Finally, to recognize and address the various social factors that influence the optimal timing of the assessment of fast changing technologies along with the need for data on health outcomes, the following recommendations by Mowatt, et. al. (1997) are offered:

- Assessment should be initiated early with a variety of approaches
- Assessment should incorporate an iterative approach
- Trials should be randomized from the outset
- Standardization of assessment methods and reporting should be developed for benchmarking and comparison
- Resource issues should be incorporated into assessments early on
- Citations and publication trends (e.g. bibliometrics) may be useful for identifying triggers for evaluation

Conclusions

Studies of telemedicine's effect on clinical outcomes continues to move forward, yet further momentum is needed to reach a tipping point for many applications within the larger health care system. Each specialty has unique clinical function within the larger clinical process that define, to a large extent its research agenda. Because of this, some fields of medicine and settings have demonstrated their benefit more easily. In general, most studies of clinical outcome in telemedicine have focused on patient satisfaction and diagnostic accuracy, with a notable lack of studies at the within the categories of clinical management and patient outcomes.

In the category of diagnosis, applications in radiology appear to be both accurate and already well integrated into mainstream health care delivery. Studies of microscopic images used in pathology are the most numerous in this review, but further study of diagnostic accuracy of interactive video applications is needed. Among specialties with diagnostic and clinical management functions, dermatology has the most evidence of diagnostic accuracy, but like pathology, more research into live interactive applications are warranted. The accuracy of echocardiograms transmitted by telemedicine appears to be well established, as are interactive applications in psychiatry. Emergency diagnostic applications are promising. Further evidence of diagnostic accuracy for ophthalmologic exam is needed.

To date, of the little evidence for the benefit of telemedicine in the area of clinical management and patient outcomes, most appears to be with home monitoring applications for CHF, diabetes, depression, and blood pressure. This is followed by dermatology and echocardiography (cardiology).

Overall, patient satisfaction appears to be well demonstrated, particularly in the field of psychiatry and dermatology, and within multi-disciplinary applications. Providers appear to be

less satisfied than patients with telemedicine alternatives, and more studies of participant satisfaction are needed for store-and-forward applications.

There are inherent limitations and significant challenges to conducting quality clinical outcomes studies in telemedicine that are either circumstantial or introduced through inadequate study designs by investigators. Several strategies to address these challenges have been offered. While it is obvious that randomized controlled trials are needed, other randomized and iterative approaches may also offer a source of important information particularly in the early phases of operations where evaluation protocols may disrupt clinical workflow. Newer research methodologies such as bibliometric studies and the use of 'tracker trials' offer additional alternatives.

Appendix

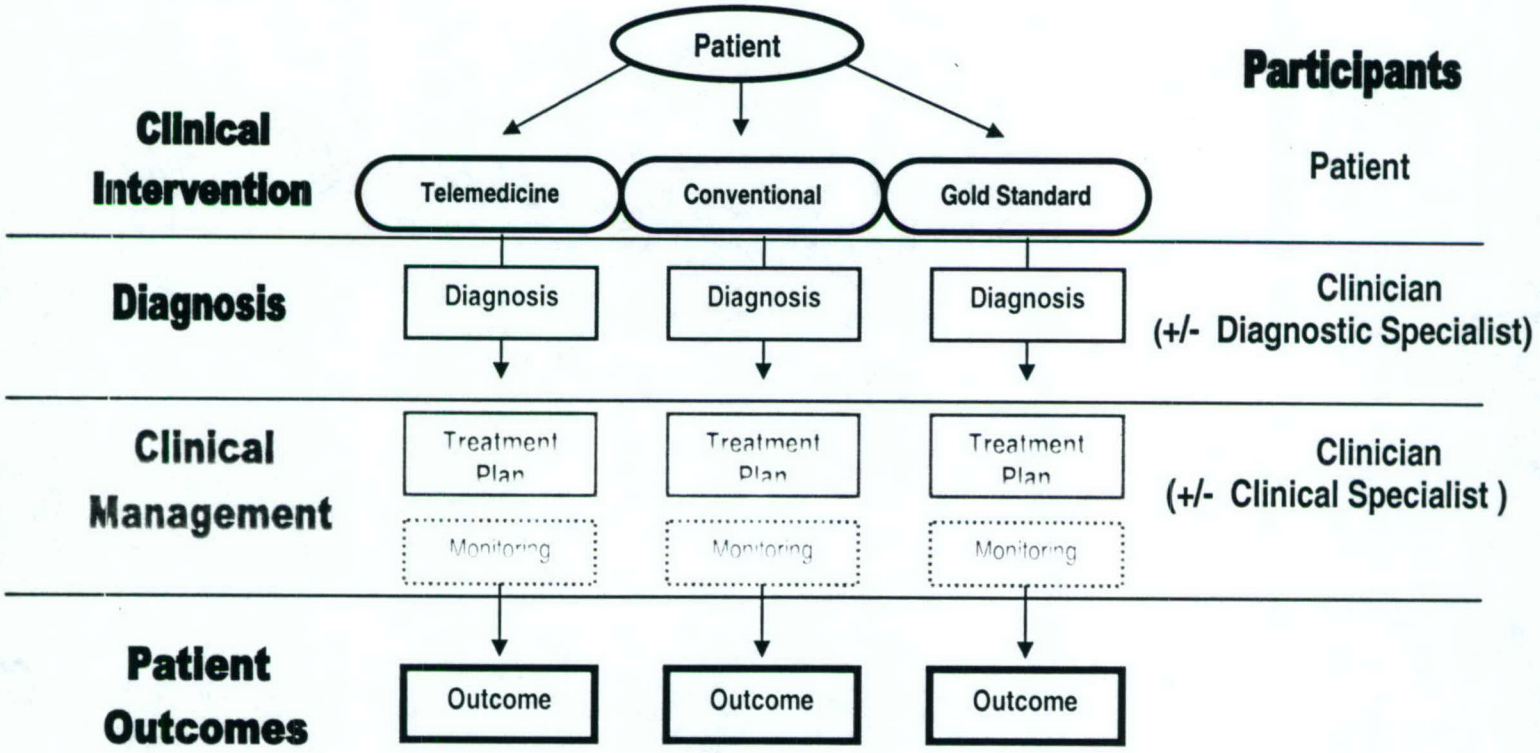


Figure 1. The clinical process and variables typically examined in telemedicine clinical outcomes studies.

Diagnostic	Diagnostic & Clinical Care	Disease monitoring
Radiology	Dermatology	Mental health follow-up
Pathology	Cardiology	Cardio-pulmonary monitoring
	Emergency Medicine	Surgery-post op
	Other subspecialties	Wound care
	Psychiatry, etc.	Disease monitoring, etc.
↓	↓	↓
Diagnosis	**Diagnosis**	
	Clinical management	**Clinical management**
	Patient outcomes	**Patient outcomes**
Participant satisfaction	**Participant satisfaction**	**Participant satisfaction**

Table 1. Expected research focus by clinical function

Table. 2 Literature Review: studies identified for analysis

Clinical Process	PubMed	Other	Total
Diagnosis:	116	34	150
Clinical Management and Patient Outcomes:	16	17	33
Patient and Provider Satisfaction	120	48	168
The total number of abstracts/articles reviewed			**351**

Table 3. Studies of Diagnosis by Specialty

Specialty	**Total**	**LI**	**SF**	**ave. n**	**supportive**	**neutral**	**unsupportive**
Radiology	20	2	18	289	17	2	1
Pathology	33	12	21	125	27	2	4
Dermatology	26	9	17	122	17	7	2
Cardiology	15	13	2	74	14	1	0
Ophthalmology	11	4	7	50	5	5	1
Psychiatry	7	7	0	25	7	0	0
Emergency Care	6	5	1	85	6	0	0
Home monitoring	5	5	0	166	4	1	0
Urology	4	0	4	33	4	0	0
Neuro	3	2	1	31	3	0	0
ENT	3	2	1	39	3	0	0
Surgery	3	3	0	60	3	0	0
Pulmonary	2	1	1	25	2	0	0
Wound care	2	1	1	234	1	1	0
Rheumatology	2	2	0	60	1	0	1
Oncology	2	2	0	74	2	0	0
Dentistry	2	1	1	20	0	1	1
Neonatology	1	0	1	31	1	0	0
Ambulatory Care	1	0	1	20	0	1	0
Microbiology	1	0	1	50	1	0	0
Neurosurgery	1	0	1	NA	0	1	0
TOTALS	**150**	**71**	**79**	**81**	**119**	**22**	**10**
Percent (rounded)		47%	53%		79%	15%	7%

Table 4. Number of Inconclusive & Unsupportive Studies: by specialty and mode

Specialty	Live Interactive		Store & Forward	
	B	C	B	C
Radiology			2	1
Pathology	1	4	1	
Dermatology	6	1	1	1
Cardiology	1			
Ophthalmology	3		2	1
Rheumatology	1			
Ambulatory Care			1	
Home monitoring	1			
Wound Care	1			
Dentistry	1			1

B = 'inconclusive', C = unsupportive'

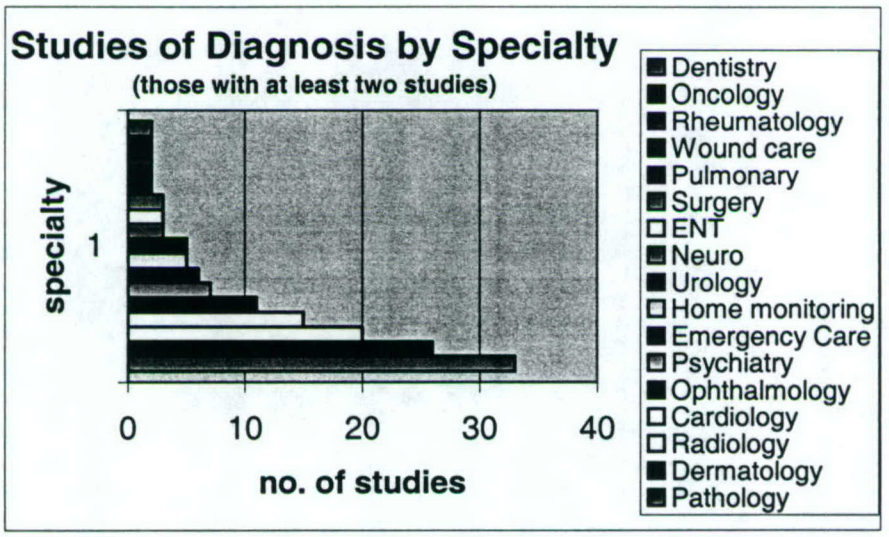


Figure 2. Studies of Diagnosis by Specialty

This page intentionally left blank.

This page intentionally left blank.

Table 5
RADIOLOGY

no.	author	year	sample	objective	result
INTERACTIVE					
1	Tachara S	2002	2133 trauma radiographs	Dx accuracy minor trauma radiographs confidence level	Overall: Accuracy 97.3%, Sens = 94.48%, Spec = 98.63% Of "sure" cases: accuracy = 98.6%, Sens = 96.72%, Spec = 99.42%
2	Franken EA Jr	1996	14 peds diseases ID'd by gen radiologist/radiograph Same 14 peds diseases ID'd by ped rad/low res digital image	Dx of pediatric radiographic images	General radiologist's Dx accuracy improved after telemed consultation with ped radiologist ROC improve from .648 to .709
STORE-AND-FORWARD					
1	Pagani, L	2003	30 pre-selected head CTs	Dx and radiologic findings; emergent head CTs	91% concordance
2	Erly WK	2003	716 head CTs	Dx of emergent head CT	95% agreement; 3% insignif disagreemnet; 2% significant disagreement
3	Kiruru MJ	2002	685 radiographs	Dx concordance of mult radiographs (vs radiologist)	phase 1: (n=446) GP selected cases- sens =.85; spec= .62 phase 2: (n=239) all consecutive cases- sens =.90; spec= .86 telerad "helped" with dx in one-third of all cases (new dx was less common)
4	Reuss L	2001	160 radiographs	Dx of known CXR abnormalities	% of abnormailites Dx: hard copy=82.3%; scanner=82.9%; digitizer=74.3%; d camera=69.7% (p<.05)
5	Zalis ME	2001 B	114 compressed CT images	Dx colonic segments CT scan at 1:1, 10:1, 20:1 compression	Sens 100% for lesions >10mm at 1:1,10:1, 20:1 Sens - 50-78% 1:1; 38-67% 10:1; 38-67% 20:1 for lesions < 10 mm (for both readers)
			114 CT images read by reader A & B	Dx concordance of colonic CT scans between readers	No significant difference (p=.3-.91)
6	Rosen MP	2001	161 laser printedus images (90 patients)	Dx accuracy of printed pelvic/abdominal US images	Of those with FU, Printed: sens 93% spec 90%; conventional: sens 95% spec 100% (no sig difference) Abdominal US: Printed: vs conventional (no sig difference)
7	Krupinski E	2000	40 bone trauma cases - digital camera	Dx concordance of bone trauma radiographs	No significant difference in Dx; Diagnostic agreement was "quite high" as was confidence ratings
8	Eng J	2000 B	120 radiograph images	Dx concordance of mult emergent radiographs	ROC .15> Rad MDs vs ER MDs; .07> Faculty vs Residents; .07 > film vs digital image Significant differences between film and digital images However equal or greater differences associated with training level and specialty
9	Zheng LM	2000	100 coronary arteries CT scans (50 are calcified)	ID coronary artery calcification	No significant difference between uncompressed images & those compressed up to 20:1 Images compressed at lower ratios were more susceptible to "blurring" Images compressed at higher ratios were more susceptible to "blocking" artifacts
10	Kroeker KI	2000	31 CT images	Dx of emergent head CT scans	84% correct; 16% incorrect; 1 clinically significant discepency; no significant difference
11	Calder LD	1999	100 mult chest and skeletal films images	Dx of mult chest and skeletal radiographs	Telerad: accuracy 91% sens 88%, spec 96%; Conventional: accuracy 90% sens 90% spec 90% Poorer accuracy with chest radiographs
12	Rosen MP	1999	80 Laserprinted images of pelvic US patients	Dx of pelvis US patients	86% agreement betwn radiologists using conventional and printed US images (10/80 "minor" discrepencies) Dx accuracy: conventional US 92%; printed image US 94% (compared to gold standard)
13	Larson A	1998	139 CXRs & skeletal radiographs with known subtle findings	ID of subtle findings	Digital: Sens in detecting nodules,pneumothoraces, interstitial disease: 58%, 75%, and 90% respectively Conven: Sens in detecting nodules,pneumothoraces, interstitial disease: 62%, 79%, and 92% respectively Digital: sens in detecting bone fractures 87%; Conven: sens in detecting bone fractures 88% None of the noted differences was statisticall significant
14	Nicolai F	1997	156 chest & abdominal radiographs	Dx accuracy of mult chest and abdominal radiographs	Dx accuracy exceeded 96%
15	Franken EA Jr	1997	153 mult ped radiographs/gen radiologist/digital image	Dx of pediatric radiographic images	Discrepensy in 13% of cases, but no "significant" difference in accuracy ROC of discrepent cases showed slightly increased accuracy for pediatric radiologists
16	Stormer J	1997 C	120 mult "difficult" radiographs (50 pre-selected)	Dx accuracy of mult "difficult" radiographs	Accuracy and sensitivity of teleradiology inferior to film radiographs
17	O'Hare	1996	? Mult tradiographs/CT scans	Dx accuracy of mult CT scans/radiographs	95% Dx accuracy of digitized radiographs & CT scans
18	Krause M	1996	446 mult high & low res film radiographs (x-ray, CT, MRI)	Dx accuracy of mult high-res radiographs, & low-res CT, MRI	Loss of Dx accuracy for high-res films, but was "satisfactory" for low res liver CT & liver MRI Average area value difference <.02

Table 6
PATHOLOGY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Moser PL	2003	270	Dx of frozen section	Control: 84.1% correct, 8.5% incorrect
2	Costello SS	2003 B	17	Dx of breast core biopsies	Ave concordance of 66.5%
3	Morgan MB	2003 C	?	Dx of routine dermatopathology specimens	kappa: .76 (telepath) vs .93 (conventional); p=.04
4	Szymas J	2001	[83] neuropath specimens	Dx concordance neuropthology paraffin sections	95% concordance for 3/3 reviewers; [100% concordance 2/3 reviewers)]
5	Dunn BE	1997	200 consec mult surgical specimens	Dx of mult surgical path slides	Clinically important concordance 99%, overall concordance 97.4% (telepath)
					Clinically important concordance 100%, overall concordance 98.5 (LM)
6	Prasse KW	1996 C	99 cytologic, 80 hematologic, 66 histologic slides	Dx concordance of cytologic slides and categorical agreement	Overall Agreement 89/99; catagorical agreement .778 to .958 (kappa .704-.948)
				Dx of hematologic slides and categorical agreement	Overall Agreement 69/80; catagorical agreement .5-1.0 (kappa .429-1.0)
				Dx of histologic slides and categorical agreement	Overall Agreement 56/66; catagorical agreement .73-.93 (kappa .47-.90)
7	Haugen OA	1999	116 frozen sections	Dx of frozen sections	90% cases were diagnosable; 3% false negatives, 0 false positives
8	Winokur TS	1998	64 frozen sections	Dx of frozen sections	Overall accuracy 96.9%
9	Callas PW	1997 C	285 surgical path cases	Dx of surgical path specimens	Telepath: 91.2% concordance [range .872-.941](95% CI)
					LM: 96.8% concordance [range .939-.985](95% CI)
					The difference between LM and telepath is significantly different (p=.009)
10	West J	1997 C	105 "difficult" cases	Dx of mult known "difficult" pathology specimens	LM was significantly better than telepath (p=.005);
					64% cases included request for review with LM
11	Dunn BE	1997	100 consec surgical specimens	Dx of routine surgical specimens	Concordance of 98.5%
12	Raab SS	1996	50 cervical-vaginal smears	Dx of routine cervical-vaginal smears	telecytology: concordance 85.6%; 34 FNs & 7 FPs
					LM: concordance 95.6%; 8 FNs & 7 FPs
					Wide range of individual performance
STORE-AND-FORWARD					
1	Lee ES	2003	50	Dx of cervical pathology	glass slides: interobserver reproducibilty .72 (first) and ..64 (second)
					digital images: interobserver reproducibilty .72 (first) and ..64 (second)
					glass slides: interobserver variation "moderate to excellent"
					digital images: interobserver variation "moderate to excellent"
2	Settakorn J	2002	100	Dx of biopsies	2 clinically significant mis-diagnoses
					"far less" turnaround time than conventional methods using postal system
3	Nehal KS	2002	[110] 50	Dx accuracy of tumor tissue; fixed tissue	Complete agreement between telepath and traditional microscope
			40	Dx accuracy presence of tumor; frozen-section	

		20 tumor frozen sections	Dx accuracy presence of tumor; frozen-section	
4 Leong FJ	2002	40 lung biopsies	Dx accuracy of lung biopsies	Bronchoscopy BXs: 2 of 20 cases were misdiagnosed; Surgical lung bx: 1 of 20 misdiagnosed
5 Cross SS	2002	100 colon polyp specimens	Dx accuracy of colon polyp histopath specimens	telepath: range .9-1.0 (excellent); significantly higher (p=.001) than glass slide: range .84-.98;
6 Singh N	2002 B	47 breast needle aspirations	Dx concordance of breast FNA smears	Absolute correlation in 80.9% of cases
7 Ito H	2001	37 kidney biopsies (31 patients)	Dx concordance of allografted kidney biopsies	30/37 cases were in agreement; insuffient/impropoer dx in 4, no dx made in 3 cases
8 Fisher SI	2001	60 hematological specimens (35 patients)	Dx concordnace of mult hematology cases	91% concordance
9 Williams BH	2001	1250 pathologic specimens	Dx mult path cases	97.7% concordance; 73.7 % absolute concordance
10 Steinberg DM	2001	10 cervical cytopathology specimens	Dx of cervical cytopathology	98.3% concordance
11 Slodkowska	2001	6 lung cancer patients	Dx lung neoplasm histo/cytopath	"high concordance"
12 Kayser	2000	54 pulmonary cytology specimens	Dx transbronchial FNA cytology	All were in general agreement; no wrong pos or negs re malig vs non-malig
13 Demichelis F	2001	210 intra-operative specimens (70 patients)	Dx of mult intraoperative frozen sections	98.6% concordance; 95.2% overall concordance
14 Leong FJ	2000	87 randomly selected breast biopsies	Dx of mult breast biopsies with known disease	98.8% concordance with microscope Dx
		44 cases of invasive tumor	Tumor typing	95.4% concordance (difference attributable to color fading - no effect on management)
		23 cases of invasive tumor	Tumor grading	91.3% concordance for carcinoma in-situ (difference has no relevance to patient management)
			Tumor grading	86.4% concordance for invasive tumor; clinically significant accuracy = 97.7%
15 Marcelo A	2000	? JPEG compressed images	Dx concordance of multiple pathology specimens	No significant differences in Dx, Image quality, confidence level, image acceptability
16 Okada DH	1999	35 skin biopsies of mult skin lesions	Dx of mult skin lesion biopsies	100% concordance
17 Ziol M	1999	100 cervical smears	Dx of cervical smears	Intraobserver agreement was "fair" to "excellent" (kappa .47-.81)
				Complete to acceptable in 68.4-85%; marked discordance in 4 cases
				No significant difference in screening slide or using videomicroscopy
18 Della Mea V	1998	299 unspecified pathology cases	Dx of unspecified path slides	"Good" diagnostic agreement
19 Weinstein MH	1997	200 consecutive prostate biosy specimens	Dx of prostate needle biopsy specimens	Set A: 77%"correct"; 9% "minor error"; 9% "major error"; 3% "deferred" (91% accuracy of non-deferred cases)
				Set B: 78%"correct"; 17% "minor error"; 1% "major error"; 4% "deferred" (99% accuracy of non-deferred cases)
20 Halliday	1997	144 mult surgical specimens	Dx of mult surgical specimens	Concordance of 88.2% of non-deferred cases (96.5% for "clinically important" Dx's); overall 74.3%
21 Weinberg DS	1996	200 routine surgical path specimens	Dx of routine surgical pathology specimens	87.5% concordance btwn LM and CD ROM Dx (Marked variability among the 4 Pathologists)
				Concodrance with consensus group: LM 95.5% and CD ROM 88.5%
				Decreased accuracy felt to be related to image selection and image quality

Table 7
DERMATOLOGY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Piccolo, D	2002	66 pigmented lesions	Dx of pigmented skin lesion	91% concordance
			43 pigmented lesions	Dx of pigmented skin lesion	77%-95% concordance
2	Gilmour	1998 C	126 patients	Dx of skin tumors	57% agreement
3	Phillips	1998 B	107 skin lesions (51 patients)	Dx agreement for skin tumors	59% agreement (kappa=.32)
4	Lesher	1998 B	115 skin complaints (60 patients)	Dx agreement of exam	78% agreement between telemed and in-person; 94% agreement between in-person examiners
5	Lowitt	1998	318 diagnoses (139 patients)	Dx agreement of exam	Paired observations: 80% agreement; 11 patients with biopsies had 7 confirmed
6	Loane MA	1998 B	427 mult skin lesions (351 pts) by RT video	Dx concordance for mult skin lesions	67% agreement between RT video & in-person exam
					9% of RT video pts had "inappropriate" plan and 20% deferred making a clinical plan
7	Oakley AM	1997 B	135 mult derm condition (104 pts) by RT video	Dx concordance for mult skin conditions	75% had correct Dx; 82% correct Dx within DDx; 12% incorrect Dx; 3% no Dx made; 4% of Dx made only in-person
8	Loane MA	1997 B	65 new mult derm clinic pts by RT camera 1	Dx concordance for mult skin conditions	59% concordance with camera 1; 76% concordance with camera 2
			65 new mult derm clinic pts by RT camera 2	Dx concordance for mult skin conditions	No Dx', wrong Dx, and missed Dx in DDx was twice as high in lower resolution camera
9	Phillips	1997 B	79 diagnoses (60 patients)	Dx agreement of exam	77.2% agreement
STORE-AND-FORWARD					
1	Lim AC	2001	53 patients with multiple skin conditions	Dx concordance of multiple skin conditions dx	mean concordance btwn 4 Dermatologists with ST telederm = 79% (73-85%)
					mean concordance btwn 4 Dermatologists with ST telederm = 86% (83-89%) when DDx included
					mean concordance btwn 11 GPs & 4 Dermatologists with ST Derm = 49%
2	Jolliffe VM	2001	? Pigmeneted lesions	Dx of pigmented lesions	No malignant tumors misdiagnosed
			in-person		
3	Barnard CM	2000	? Multiple skin diseases	Dx of multiple skin diseases	
4	High WA	2000	92 patients with 106 derm complaints	Dx accuracy in pts with mult derm complaints	81-89% concrdance for all 3 Dermatologists (4-8% clinically-relevant disagreement)
					88-100% concordance in"high confidence" cases; 84-98% concordance in "high quality image" cases
					Dx was comparable between disease types (except benign neoplams 22-46%)
					Chi square : signicant difference in accurate Dx btwn "high confidence' cases and "low confidence" cases (P<.05)
					Chi square: significant difference n Dx accuracy btwn "poor" & "adequate" images (2 Dermatologists)
5	Braun	2000	51 patients with 55 pigmented lesions	Dx accuracy of pigmented lesions	Dx accuracy was superior for Dx of malig melanocytic lesions
6	Loane MA	2000 C	96 multiple skin lesions (SF)	Dx concordance of multiple skin lesions	51% concordance for Dx
					44% concordance for Tx
7	Taylor P	2000	190 patients	Dx agreement 13 later by telemed	77% agreement
8	Krupinski EA	1999	308 mult derm clinic patients/ dig camera images/derm panel	Dx concordance of mult derm clinic patients	83% concordance between digital camera images & in-person exam Dx's;
					INTRAdermatologist concordance 84%; INTERdermatologist concordance 81%
					Rated Confidence: "very definite" to "definite" in 62%
					Concordance 72% with those cases that had biopsy as gold standard

					Image "sharpness" and "color quality" "good" to "excellent" in 83% & 93% respectively
9	Piccolo D	1999	66 pigmented lesion images received by email/Austria	Dx concordance for pigmented lesions	91% concordance, not significantly different (Wilcoxon test p=.10)
					Dx accuracy was not related to image quality, but depended on the level of "diagnostic difficulty" (Spearman p=.01)
10	Whited JD	1999 B	168 mult skin lesions imaged dermatologists	Dx accuracy for mult skin lesions	Dx concordance among clinic-based dermatologists for single most likely Dx = .54 (95% CI .46-.61)
					This increased to .92 when DDx was included
					Results for teledermatologists were comparably reliable to in-person
11	Lewis K	1999	141 patients	Dx agreement for skin tumors	Sensitivity 88%, specificity 80%
12	Whited JD	1998	12 pts 13 ? skin CA lesions / 2 dermatologists/ dig image	DDx agreement of suspected skin CA	"almost complete agreement" on differential Dx and biopsy recommendations
13	Kittler H	1998	50 images of pigmented lesion/ digital	Dx accuracy of pigmented lesions by epiluminescence	Mean area under the curve .81 for photo slides AND digital images
14	Harrison PV	1998	? photos of mult dermlesions/ photo/dermatologist	Dx accuracy of photographs of mult derm lesions	Dermatologist c photograph = =71%; GP in-person = 49%
					malignancy detected in 94% of cases by dermaologists c photographs & 70% by GP in-person
15	Zelickson BD	1997	30 skin conditions (29 patients)	Dx agreement	67% correct Dx with history alone; 85% with image alone; 88% with history and image
16	Lyon CC	1997	100 patients	Dx agreement between in-person trainee & remote MD	10% disagreement in patients with rashes; 4% disagreement in patients with tuomors
17	Kvedar JC	1997	116 pts with mult skin conditions by digital image	Dx concordance for mult skin conditions	61-64% concordance; this increased to over 75% when clinician certainty and image quality were high
					Diagnostic certainty had the highest impact on agreement; image quality had a modest impact

Table 8
CARDIOLOGY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Belmont, JM	2003 B	76 pediatric patients	Detection of presence/absence of heart disease (telestehoscope)	kappa=.63
				Dx certain heart sounds	kappa=.65-.75
				Dx certain heart sounds	kappa < .60
			49 pediatric patients	Dx certain heart sounds	kappa = 90, agreement = 96%, sens =94%, spec = 100%
2	Scalvini S	2002	? adult patients with chest pain	Dx of chest pain of cardiac origin	sens = 97.4%; spec = 89.5%, diagnostic accuracy = 86.9%
3	Sable CA	2002		Dx of cardiac abnormailities (vs video-taped study)	1 minor change in Dx; 3/264 changes on FU in-person ECHO
4	Sable C	1999	60 ECHOs mult crit ill neonates by RT video	Dx accuracy of critically ill neonatal ECHOs	100% concordance of neonatal ECHOs
5	Mulholland HC	1999	63 ECHOs neonates with ? Cardiac defects by RT video	Dx accuracy of neonatal ECHOs with RT transmission	3% not able to Dx; 93% Dx concordance by subsequant in-person ECHO
6	Tobis J	1999	40 cardiac cath cases by RT video	Measurement of angiographic images	Measurements of angiographic images were "very close";
					No differences in transmitted or in-person ECHO
7	Finley JP	1997	26 pediatric pts of various age ECHOs by RT	Dx concordance of pediatric ECHOs various ages	No important discrepancies in pediatric ECHOs
8	Afset JE	1996	38 cardiac patients by ECHO by RT	Dx concordance of routine ECHOs	No systematic measurement erros in routine ECHOs; No clinically sig differences in Doppler quantification
					No important M-mode discrepensies, except mised LVH in 6pts; 2-D exam missed 3 clinically significant findings
9	Trippi JA	1997	163 ED pts c chest pain	Dx sens/spec of dobutamine stress tele-ECHO for chest pain pts	sens = 89.5%, spec = 88.9% for dobutamine stress tele-ECHO; negative predictive value of 98.5%,
10	Trippi JA	1996	187 emergent ECHOs on mult ED patients by RT	Dx accuracy of emergent ECHO for ED patients	99% concordance of tele-ECHO with in-person; 96.3% concordance for "serious" abnormailities
11	Casey F	1996	9 pediatric patients	Dx concordance of exam	88% correct diagnose
12	Belmont	1995 B	116 pediatric cardiac patients	Dx concordance of stethoscope exam	Kappa coefficient = .64 between acoustic and tele-stethoscope
13	McConnell	1996	21 pediatric patients	Dx agreement for evaluation of murmurs	90% agreement

STORE-AND-FORWARD

1	Cotton JL	2002		Dx of heart structures and function	no significant difference
2	Srikanthan VS	1997	112 ED pts c suspected MI	Agreement re use of thrombolytics	8 pts saved from unecessary use of thrombolytics, 4 received it when otherwise they would not have

Table 9
SURGERY

no.	author	year	sample	objectives	results
INTERACTIVE					
1	Wagner A	2002	7 TMJ surgery cases	Accuracy of digitizing probe measurement in TMJ punctures	ave error of digitizing probe = 1.3 mm; ave SD = .7mm
2	Demartines N	2000	112 pts undergoing endocrine surgery	Dx with CT/MRI	
				ID of fine organ structure in GI/Endo surgical patients	RT telemed 89.3% correctly identifed, in-person CT/MRI 95.5% correctly identified
				ID pathological findingin GI/Endo surgical patients	RT telemed 98.2% correctly identified, in-person CT/MRI 99.1% corectly identified
				Ability to Dx in GI/Endo surgical patients	RT telemed 84.4% made Dx, in-person CT/MRI 93.3% made Dx
3	Demartines	2000	60 randomly selected surgical cases	Ability to ID target organ in multiple presented surgical cases	100% ID target organ
				Ability to ID organ structure & pathology in multiple surgical cases	93.3 % organ structure "well defined"
				Ability to ID fine structure in multiple surgical cases in multiple surgical cases	58.3% fine structure "well defined"
				Ability to Dx accurately in multiple surgical cases	28.3% "accurate" Dx; 41.7% "probable" Dx; 26.7% "possile but uncertain"; 3.3% "not possible"
STORE-AND-FORWARD					
1	West J	1997	? CT/MRI/PET scans of neurosurgical pts	Taget registartion errors in neurosurgery	(statistics in article)

Table 10
EMERGENCY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Berger J	2001	? ER patients needing consultation	? (unclear)	?
2	Tachakra S	2000	200 patient consultations (minor trauma)	ID color change in multiple trauma patients	97% concordance
				ID swelling or deformity in multiple trauma patients	98% concordance
				ID decreased joint movement in multiple trauma patients	95% concordance
				ID presence of tenderness in multiple trauma patients	97% concordance
				Assess WT-bearing in multiple trauma patients	99% concordance
				Estimate of severity of illness in multiple trauma patients	1 overestimated, 5 underestimated
3	Lee	1998	123 studies/460 radiographs (90 patients)	Dx agreement of exam	Initial diagnosis changed in 30% of cases
4	Kofos	1998	15 pediatric patients	Dx agreement exam	Sensitivity and Specificity of abnormal findings: 87.5% and 93% respectively

5	Tachakra S	1998	150 patients	improvement in diagnosis	Telemed improved sensitivity from 90% to 97%, sensitivity from 96% to 99%

STORE-AND-FORWARD

1	Jacobs MJ	2002	20 patient radiographs with maxofacial trauma	Dx of maxifacial fracture	plain radiography was more accurate in assessing position of fractures Dx of Oral & Maxofacial surgeons using teleradiology was broadly comparable to EM physicians using plain radiography

Table 11
OPHTHALMOLOGY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Bowman, RJ	2003 B	80 pts with emergent eye complaints	Dx concordance of emergent eye problems	Control: 75% complete concordance, 5% impt disagreement Case(torchlight): 40% complete concordance, 10% imt disagreement;(chi(2) =10.7, P=.007) Case(slitlamp): 58% complete concordance, 5% imt disagreement;
2	Threlkeld AB	1999 B	50 eyes [25 patients] by slit-lamp exam for mult probs	Dx accuracy of slit-lamp exam for eye & lid diseases	Sens%/Spec%: eyelid mass=100/64, conjunctival pigment=100/85, post synechiae=100/100, blepharitis=80/0 (Cont) iridotomy= 83/93, pinguecula=70/93, iris lesions=75/93, corneal scar=56/98, chamber inflammation=0/100 (cont) nuclear cataract=57/93, intraocular lens presence=37/100
3	Marcus DM	1998 B	17 HIV + pts undergoing ophthalmoscope exam	Dx accuracy in detecting HIV-retinopathy	HIV pts: disease-free eyes-spec=95%; HIV retinopathy sens=83% (Dx accuracy =100% in eyes s cataracts)
			20 diabetic pts undergoing ophthalmoscope exam	Dx accuracy in detecting DM diease	DM pts: Unable to ID 46% glaucoma pts & 36% DM retinopathy - assoc c pts that have cataracts (cont) sens for DM retinopathy, glaucoma, and cataract was poor (29%, 50%, 41% respectively)
4	Nitzkin JL	1997	58 patients	Dx agreement	85.9% agreement in diagnois
STORE-AND-FORWARD					
1	Cook HL	2000	11 retinal imaging cases	ID of retinal pathological signs in multiple known pathologic cases	Total ratings scores higher for transparencies vs color or B&W digital images (*p=.00006) No significant differences in transparency vs color digital images (p=.09) or monochrome (p=.11) Experts were better than trainees at detecting pathology
2	Yogesan K	2000 C	49 eyes (25 patients)	Dx accuracy (dig camera vs fundus camera)	Kappa < .3
3	Liesenfeld B	2000	129 diabetic patients	Dx agreement (dig camera vs fundus photo)	Sensitivity 85%, specificity 90%
4	Yogesan K	2000 B	51 eyes (29 glaucoma screened patients)	Dx agreement for glaucoma	Sensitivity .68-.79, specificity .67-.87
5	Schwartz SD	2000	19 eyes (10 patients)	Dx agreement for retinopathy of pre-maturity	89-95% agreement
6	Li HK	1999	26 eyes 32 patients	Dx agreement for glaucoma among optometrists	80.8% agreement in cup-to-disc ratio
7	Briggs R	1998 B	90 mult retinal images-digitized	Dx accuracy of retinal images c mult conditions	Significant difference in reported confidence, lower for those viewing the digital image

Table 12
OBSTETRICS/GYNECOLOGY

no.	author	year	sample	objective	result
INTERACTIVE					
1	Ferris DG	2002	? patients undergoing colposcopy exam	Dx concordance colposcopy exam (vs histology report)	Concordance 59.7% (kappa = .31) for local colp, 52.7% (kappa = .22) for site experts
					Concordance 55.7% (kappa = .27) for distant experts with RT video
					Concordance 49.7% (kappa = .16) for distant experts with SF video
2	Allen-Davis JT	2002	253 patients with vaginal complaints	Dx of vulvaginal complaint	poor agreement between nurse and practitioners(MD and PA)
					Practitioners (MD & PA) failed to acurately dx vaginitis
STORE-AND-FORWARD					
1	Wooton R	1997	192 viewings (18 pts) of prenatal US at varied bandwidth	Dx accuracy of pre-natal US's viewed at varying bandwidth	(1920 kbit/s) 85% Dx'd correctly or 'half correctly'; 384 kbit/s 69% Dx'd correctly or 'half correctly' (significant difference)

Table 13
HOME MONITORING
INTERACTIVE

no.	author	year	sample	objective	results
1	Moller, DS	2003	411 patients with HTN	Accuracy of home telemed BP measurement	Correlation clinic vs ambulatory BP "weak" (sys r=.499, dia r = .543)
					Correlation home tele vs ambulatory BP "strong" (sys r=.847, dia r = .812)
					Stronger correlations btn home tele BP & Amb BP at noon and afternon
2	Moller DS	2003	362 patients with HTN	Accuracy of home bp (tele) and clinical bp vs ambulattory BP in pts with HTN	Correlation btwn CBP and ABP was weak [sys :r =.343, dias: r = .430]
					Correlation btwn HBP and ABP was strong [sys :r =.804, dias: r = .776]
					Strength of the correlation of HBP and ABP increased over 5-day period, and with afternoon readings
3	Aris	2001	4 patients with HTN	Accuracy of remote BP measuement	Measurement error: systolic 1.7-2.7%, diastolic 2.7-3.2%
4	de Lusignan	2000	20 patients	Accuracy of vital signs	Within confidence interval for heart rate; outside interval for resp rate, temp, BP
5	Finkelstein J	2000	31 patients	Accuacy of spirometry	No difference between supervised and unsupervised home spirometry

Table 14
RHEUMATOLOGY

No.	author	year	sample	objective	results
INTERACTIVE					
1	Leggett P	2001	100 pts with mult rheum probs	Dx of mult rheumatologic conditions (vs traditional)	97% concordance
2	Graham LE	2000 C	20 pts with mult rheum probs	Dx accuracy in pts with mult Rhematologic complaints	35% concordance by telephone; 40% concordance by video conferencing
					85% of small joint swellings could not be identified
					important clinical signs could not be visualized

Table 15
ENT

INTERACTIVE

1	Ullah R	2002	42 patients in need of ENT consultation	Dx agreement of ENT conditions	34/43 had correct diagnosis
2	Sclafani AP	1999	45 patients	Dx agreement of ENT conditions	85% agreement for non-interactive video; 64% for store-and-forward exam

STORE-AND-FORWARD

1	Furukawa M	1998	29 patients	Dx agreement of laryngoscopy	100% agreement among three ENT specialists (text and images)

Table 16
Psychiatry

INTERACTIVE

1	Shores, MM	2004	16 psychiatry patients	Dx of dementia	100% concordance
2	Elford R	2000	23 psychiatry patients	Dx agreement	96% agreement
3	Kirkwood KT	2000	26 patients	Neuropsychiatry testing agreement	All results within 95% confidence interval of difference
4	Ruskin PE	1998	30 patients	Dx agreement	Kappa nearly identical for in-person vs in-person, and in-person vs remote (.70 to 1.0)
5	Ball C	1998	11 patients	Mini mental status exam (MMSE) agreement	Pearson correlation coefficient was .89
6	Baigent MF	1997	63 patients	Dx agreement	Kappa =.85 for in-person vs. in-person; kappa = .70 telemed vs. in-person
7	Montani C	1996	9 elderly patients	Reliability of MMSE and complex figure test (CFT) rating	Small but significant difference for remote patients

Table 17
Pulmonology

INTERACTIVE

1	Pacht ER	1998	40 patients	Accuracy of history and physical exam	Kappa varied .29 (history) to .66 (physical exam)

STORE-AND-FORWARD

2	Abboud S	1996	10 healthy individuals	Dx accuracy, reliability spirometry in healthy pts	Spiromtry results FVC and FEV1 were within 5% of each other 90% of the time

Table 18
Ambulatory care

STORE-AND-FORWARD

1	Houston MS	1999	B 20 patients	Dx agreement	Complete agreement in all cases

Table 19
Neonatology

STORE-AND-FORWARD

1	Yamamoto LG	1996	31 neonatal patients	Dx accuracy of CXRs between GPs and Neonatologists	Neonatologists ID'd 91-98% of findings; pediatricians ID'd 82% of findings

Table 20
Dentistry

INTERACTIVE

1	Baur DA	1998	B 13 randomly selected pts for oral exam	Dx accuracy of telemed orthognantic exam in random pts	Absolute value of within-subject difference between the 2 exam methods was significantly different

STORE-AND-FORWARD

1	Patterson S	1998	C 27 dental patients	Dx agreement	Kappa for agreement on tooth decay = .50-.58

Table 21
NEUROLOGY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Craig JJ	1999	23 patients	Dx agreement between jr physician (telemed) and in-person neurologist	Kappa varied across findings from .42 to .1.0
2	Hubble JP	1998	9 patients with Parkinson's disease	Functional status assessment	High correlation for subjective (.97), objective (.91), and staging (.88)
STORE-AND-FORWARD					
1	Johnston, KC	2003	60 pts needing CT eval	Dx of acute stroke by CT	kappa=1, sens & spec = 100% Overall: No significant difference between digital and plain film radiographs

Table 22
UROLOGY

no.	author	year	sample	objective	results
STORE-AND-FORWARD					
1	Kuo RL	1999	14 mult urologic cases with digitized images & histories	Dx accuracy in mult urologic radiologic studies	Dx accuracy ranged from 85.7% to 100% (differences between attending & resident physicians was not significant)
2	Lee BR	1999	94 IVPs, 4 cystourethrograms, 2 nephrostograms	Dx accuracy of IVPs/cystourethrograms/nephrostograms	88% concordance btwn urologists using telerad images and radiologist using orig radiograph
3	O'Sullivan	1997	50 abdominal x-rays- pts with ? urinary calculi/ gp A 50 abdominal x-rays- pts with ? urinary calculi/ rgp B	Dx accuracy in abdominal urologic radiologic studies	Plain films: Dx accuracy for all urologists 86.5%; Digital images: 81.5% (p>.2) No difference with respect to iamge quality, "diagnostic difficulty", reader confidence
4	Averch TD	1997	26 abdominal x-rays- pts with ? urinary calculi/ 25 abdominal compressed x-rays- pts c? urinary calculi/	Dx accuracy in abdominal urologic radiologic studies Dx accuracy in compressed abdominal urologic radiologic studies	Attendings: Dx accuracy 96%; Fellows: Dx accuracy 92% (differences between attending & fellows was not significant) Attendings: Dx accuracy 96%; Fellows: Dx accuracy 84% (differences between attending & fellows was significant) Overall: No significant difference between digital and plain film radiographs

Table 23
WOUND CARE

no.	author	year	sample	objective	results
INTERACTIVE					
1	Kim HM	2003 B	430 wound assessments (70 patients)	Wound status agreement	% agreement ranged from 67.1- 88.8 sens range .32 - .63; spe ranged .80 -.91
STORE-AND-FORWARD					
1	Wirthlin DJ	1998	38 wounds (24 patients)	Dx/assessment agreement	Sensitivity .86, specificity .72

Table 24
ONCOLOGY

no.	author	year	sample	objective	results
INTERACTIVE					
1	Stalfors	2001	98 oncology patients	Dx with TNM classification & Tx for head/neck CA	91% concordance sens range .32 - .63; spe ranged .80 -.91
2	Hashimoto	2001	50 patients	Treatment center verification by CT with reconstructed images	No statistical difference

Table 25
Neurosurgery

STORE-AND-FORWARD					
1	West J	1997 B	? CT/MRI/PET scan	Dx accuracy of target registration	'Satisfactory' results

Table 26
Microbiology
STORE-AND-FORWARD

1	McLaughlin WJ	1997	50 gram stain slides	Dx accuracy	No staistic difference (LM =95.4%, digital image = 95.3%)

Table 27 Role of experience

1	Krupinski EA	1996 ?	radiographs viewed as digital images	Effect of experience on Dx performance	Positive correlation between years of experience & performance for conventional practice

Specialty-Setting	% of total	Studies Total	Store-and-Forward	Interactive
Home-based monitoring	39.4%	13	0	13
Dermatology	15.2%	5	3	2
Multi-specialty consults	12.1%	4	2	2
Emergency	6.1%	2	0	2
Psychiatry	6.1%	2	0	2
Cardiology	6.1%	2	0	2
Ophthalmology	3.0%	1	1	0
Wound care	3.0%	1	1	0
Urology	3.0%	1	0	1
NICU	3.0%	1	0	1
Dentistry	3.0%	1	1	0
TOTAL		33	8	25
Percent			24.2%	75.8%

Table 28. Studies examining Clinical Management and Patient Outcome

This page intentionally left blank.

Table 29. Studies examining Clinical Management and Patient Outcome

INTERACTIVE

Author	Year	Specialty	Sample	Objective(s)	Outcome (* = 'inconclusive' rating)
Mullen B	2003	Home care: lung transplant candidates	119 patients	Compare to phone calls	No significant difference in clinical outcomes
Chan DS	2003	Home care: pediatric asthma monitoring	10 patients	Compare to office-based education and monitoring	Improved inhaler technique scores and peak flow values No change in QOL per patients, but increased per caregiver
Bellazi R	2003	Home care: DM management	67 patients	Compare HgbAc as indicator of DM control	Significantly lower HgbA1c than controls
Ryan P	2003	Home care: chronic medical & mental illness	791 chr ill 120 ment ill	Compare to historical conventional care	Improved perception of functional status (e.g. pain physical and social)
Benatar D*	2003	Home care: CHF monitoring	216 patients	Compare to visiting RN	At 3 months: Fewer re-admissions ($p<0.001$); shorter length of stay ($p<0.001$); decreased charges ($p<0.02$)
LaFramboise LM	2003	Home care: CHF monitoring	60 CHF patients	Compare to conventional home care (RN visits and telephone)	No differences in functional status, depression, or health-related quality of life.
Hunkeler EM	2000	Home care: Depression care	302 patients on antidepressants	Compare telephone-based telehealth to usual care	At 6 months – improved Hamilton Depression rating (57% vs 38% p=.003), Beck Depression Inventory (48% vs 37% p=.05). Improved mental functioning & satisfaction at 6 weeks. No difference from telephone peer support
Cherry JC	2003	Home care: complex illness	? patients	Compare utilization and clinical impact to pre-intervention	Decreased inpatient admissions, ER visits, hospital stays, increased quality of life
Myer M	2002	Home care: Chronic disease	791 patients	Compare to a comparison control group	Decreased hospital admissions by 63%; reduction in hospital bed stays 60%; 44% decrease in ER visits; 64% decrease in VA nursing home admissions; 88% decrease in nursing home bed stays
Cherry JC	2002	Home care: DM management	? patients	Compare to historical conventional care	Reduced utilization and charges, improved QOL $747 savings per pt/yr; 32% less inpatient admissions ($p<0.07$); 44% less ER visits ($p<0.28$); 49% less outpatient visits ($p<0.001$)
Johnston B	2000	Home care: chronic illness management	212 patients	Compare to conventional home care alone (RN visits and telephone)	No difference in quality indicators (medication compliance, knowledge of disease, ability for self-care), satisfaction, or service utilization.
Bruderman I	1997	Home care: asthma monitoring	39 patients	Compare tele-spirometry vs. conventional emergent care	Decision to dispatch emergency mobile unit correlated with spirometry on 56% of occasions *
Friedman RH	1996	Home care: BP monitoring (computer-controlled telephone system)	267 patients	Compare to those using conventional care	Weekly use of computer-controlled telephone system improved medication adherence ($p=.03$) and mean diastolic BP control ($p=.02$)
Dimmick SL	2003	Multiple health & behavioral services (10)	? patients	Compare to conventional care	Decreased CHF hospitalizations; improved glucose control for diabetics
Wallace P	2002	Multi-specialist consultations	2094 patients	Compare to in-person exam outcomes	(pending)
Rendina MC	1998	Neonatal ICU	87 infants	Compare to convention of mailing video-taped ECHO	Reduction in NICU length of stay was reduced by 5.4 days (non-significant)
Belmont JM	1995	Cardiology	116	Compare tele-stethoscope to conventional	Sensitivity and Specificity of need for follow-up or ECHO; 88% and 97% respectively
Srikanthan VS	1997	Cardiology	112	Compare use of faxed EKG to conventional care	Agreement to use thrombolytics in 87.5% of cases
Ellis DG	2001	Correctional: ER consultation	530 ER care records	Compare to in-person ER visit	81 of 126 teleconsultations (64%) avoided transportation to the ER Ave time 30 minutes vs 2.75 hours for ER visit
Lee JK	1998	ER	460 radiographs (90 patients)	Compare teleradiology to in-person exam	Initial treatment plan was changed for 26% of patients
Loane MA	1998	Dermatology	351 patients (427 diagnoses)	Compare video to in-person exam	Same management plan in 64% *
Phillips CM	1998	Dermatology	107 lesion (51 patients)	Compare interactive video to in-person exam	Recommendation for biopsy 86% agreement (kappa=.47) *

Zaylor C	1999	Psychiatry	49	Compare to in-person consultation Global Assessment Functioning (GAF) scores	No significant difference in GAF scores. Telemed patients had greater attendance, and shorter visits.
Elford	2000	Psyhiatry	23 patients	Compare telepsychiatry to in-person	Management plan agreement in 96% of patients
Hays WS	1998	Urology	32 patients (18 are simulated)	Compare teleconsultation to in-person	Treatment changed in 50% of actual patients and 17% of simulated patients *

STORE-AND FORWARD

Author	Year	Specialty	Sample	Objective	Outcome
Kedar I	2003	Multispecialty/ Oncology	71 patients (90% cancer patients)	Compared to conventional care	90% had new treatment recommendations Turn around time decreased from 19 to 6.8 days
Zelickson BD	1997	Dermatology	29 patients (30 conditions)	Compare teleconsultation to in-person	90% agreement in management with history and images 87% agreement with images alone 70% agreement with history alone
Whited JD	1999	Dermatology	129 patients (168 lesions)	Compare teleconsultation to in-person	77% agreement on management among clinical examiners 56-77% agreement between in-person and teleconsultation 64-83% agreement between teleconsultants
Taylor P	2000	Dermatology	190 patients	Compare teleconsultation 13 months after in-person	31% of cases could have been managed by GP
Schwartz SD	2000	Ophthalmology	10 patients (19 eyes)	Compare teleconsultation to conventional	42% of cases had agreement of management plans
Wirthlin DJ	1998	Wound care	24 patients (38 wounds)	Compare photograph to in-person exam	.87 sensitivity and .78 specificity in wound management decisions
Houston MS	1999	Multi-specialty/ Outpatient	20 patients	Compare store and forward consultation to in-person	79% of store-and-forward cases had additional treatment recommendations; 93% of in-person cases had additional treatment recommendations *
Patterson S	1998	Dentistry	27 patients	Compare teleconsultion to in-person oral exam and management	Agreement for extraction and filing (kappa=.93-1.0)

Figure 3. Patient and Provider Satisfaction: Specialties

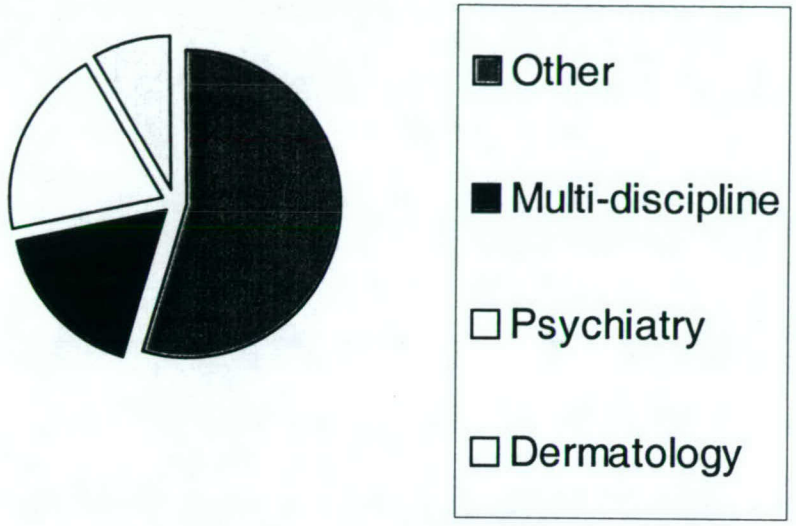


Figure 4.
Patient and Provider Satisfaction: Setting

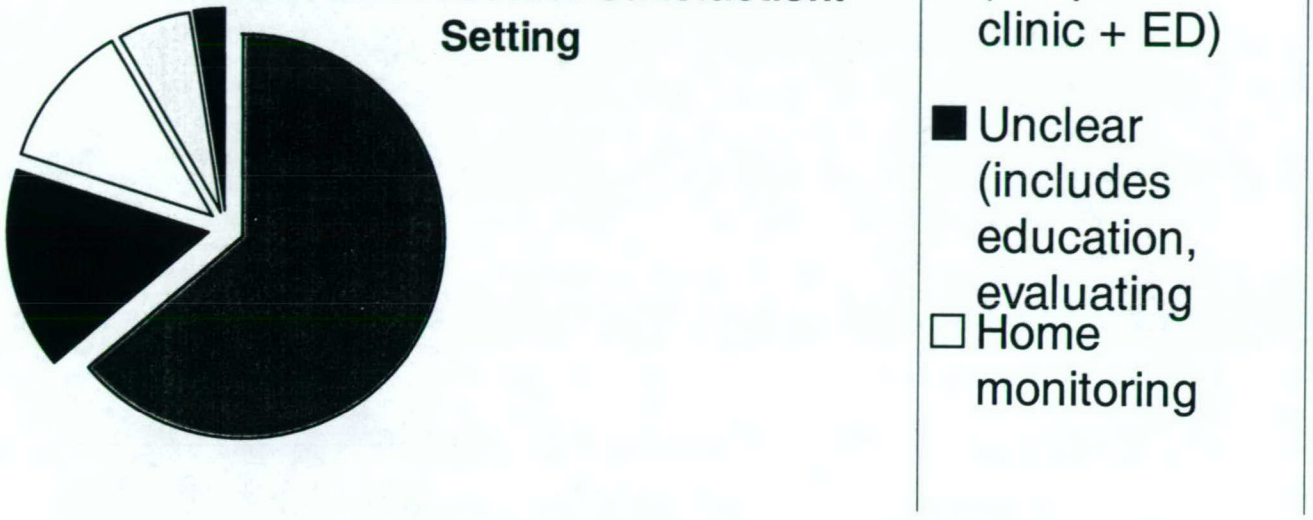


Table 30 . Classification of Study Design
(adaptation of Jovell and Navarro-Rubio's classification (Roine 2001)

1. Meta-analyses of randomized controlled trials (RCTs)
2. Large-sample RCTs
3. Small sample RCTs
4. Non-randomized controlled prospective studies
5. Non-randomized controlled retrospective trials
6. Cohort studies
7. Case-control studies
8. Noncontrolled clinical series, descriptive studies, consensus methods
9. Case reports and anecdotes

References (Note: journal citations of the 351 studies evaluated in the review are pending)

Aoki N, Dunn K, Johnson-Troop KA, Turley JP. Outcomes and methods in telemedicine evaluation. *Telemedicine Journal and e-Health.* 2003 9(4):393-401.

Bashshur RL: Where are we in telemedicine/telehealth, and where we go from here. *Telemedicine Journal and e-Health*, 4,(1998)1-4.

Berg M. Patient care information systems and health care work: a sociotechnical approach. International Journal of Medical Informatics 55 (1999)87-101

Field MJ (Ed): (1996) *Telemedicine: A guide to assessing telecommunications in health care.* Institute of Medicine Washington, D.C.: National Academy Press.

Finch TL, May C, Mair SM, Mort M, Gask L. Integrating service development with evaluation in telehealthcare: an ethnographic study. British Medical Journal 2003 327:22:1-5

Grigsby J, Rigby M, Hiemstra A, House M, Ollsson S, Whitten P. Telemedicine/Telehealth: An international perspective. The diffusion of telemedicine. Telemedicine Journal and e-Health, 2002;8;79-94.

Hailey D, Roine R, Ohinmaa A. Sytematic review of evidence for the benefits of telemedicine. Journal of Telemedicine and Telecare 2002;8(Suppl.1):S1:1-7

Hersh W, Helfand M, Wallace J, Kraemer D, Patterson P, Shapiro S, Greenlick MA. A systematic review of the efficacy of telemedicine for making diagnostic and management decisions. *Journal of Telemedicine and Telecare* 2002;8:197-209

Holle R, Zahlmann G. Evaluation of telemedical services. IEEE Transactions on information technology in biomedicine. 1999;3;83-89.

Lilford JG, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. *British Medical Journal* 2000;320:43-6

Mowatt G, Bower, DJ, Brebner JA, Cairns JA, Grant AM, McKee LM. When and how to assess fast-changing technologies: a comparative study of medical applications of four generic technologies. *Health Technology Assessment* 1997;1:14

Roine R, Ohimaa A, Hailey D. Assessing telemedicine: a systematic review of the literature. *Canadian Medical Association Journal* 2001;165(6):756-71

Williams T, May C, Mair F, Mort M, Gask L. Normative models of health technology assessment and the social production of evidence about telehealth care. *Health Policy* 64 (2003) 39-54.

Yellowlees P. Practical evaluation of telemedicine systems in the real world. *Journal of Telemedicine and Telecare.* 1998; 4 Suppl 1:56-57.

Health Services Research and the Evaluation of Telemedicine

Jim Grigsby
Angela G. Brega
Patricia A. DeVore

Division of Health Care Policy and Research,
Department of Medicine,
University of Colorado Health Sciences Center
Denver, CO

Running Head: Health Services Research and Telemedicine

Address Correspondence to:
Jim Grigsby, PhD
University of Colorado Health Sciences Center
13611 East Colfax Avenue, #100
Aurora, CO 80011-5704

Introduction

Telemedicine is the use of telecommunications and information technology to provide health services to persons who are at some distance from a provider. Some prefer the term *telehealth,* which they consider to include a broader range of providers and applications, using the term *telemedicine* to refer to services provided by physicians. The concept of telemedicine, or telehealth, encompasses a very wide range of technologies and applications. The former includes the use of the telephone, radio, fax, video conference equipment, computer, Internet, World-Wide Web, specialized microprocessor devices for the capture and transmission of physiologic and other clinical data, virtual reality, and telerobotic tools.

At its most fundamental level, telemedicine is a means of delivering care when the provider and patient are not physically present in the same location. The range of possible clinical applications of telemedicine, therefore, is as wide as all of health care itself. Consequently, discussions of such issues as the effectiveness, efficacy, or cost of telemedicine must be constrained to specific health care niches. For example, laparascopic surgery, psychiatric assessment, retinal photography, and home health care are very different from one another in their aims, practitioners, and component processes. All can be accomplished using telemedicine technology, but any attempt to make general statements about the effectiveness of telemedicine across such disparate health services and settings is necessarily doomed to failure.

Brief History of Telemedicine Research

Telemedicine had its beginnings in the United States in the 1950s as a means of providing psychiatric consultation using interactive video (IAV) to a rural state psychiatric hospital (Benschoter, 1967). The quality and capability of the costly technologies then available was prohibitive, and this method of providing health services expanded very slowly. By the early 1990s, improvements in telecommunications and computer technology made the use of telecommunications and information technology for remote health care delivery more affordable and reliable. Interest in telemedicine surged.

Over the past 10 or 12 years, telemedicine technology has undergone considerable development. The cost of equipment and of access to wide bandwidth communications channels has declined significantly, a wide range of applications has been devised and implemented, and the quality of the published literature on telemedicine has improved. Yet despite the fact that telemedicine has been practiced for nearly 50 years (Benschoter, 1967), and longer if one considers the use of the telephone in the delivery of health care services, the subject has not been researched adequately.

Despite the growth of telemedicine, limited research exists to support its medical- and cost-effectiveness (Grigsby et al., 1995; Strode et al., 1999). Studies conducted in the 1970s showed that IAV telemedicine, audio-only telemedicine, and in-person care were roughly comparable in quality (Conrath et al., 1977; Moore et al., 1975), but few studies conducted in recent years have been as well designed, or have had as many subjects, as this early work. Nevertheless, data exist to demonstrate the benefits of telemedicine (Hailey et al., 2002). In large part, the relative paucity of research reflects the fact that telemedicine has proliferated slowly; few providers have adopted telemedicine as a routine means of providing care, and the volume of patients examined and treated using the technology has remained low.

Although a number of variables have served to limit the rate of diffusion of telemedicine (e.g., Grigsby et al., 2002), a lack of data on its effects has contributed to a sense of uncertainty on the part of individuals and organizations that might otherwise adopt it for use in the clinical setting uncertainty about whether it will work, whether providers and patients will use it, and

what it will cost. Such uncertainty has been identified as a major factor in limiting the diffusion of innovations (Rogers, 1995).

Purpose of this Paper

The objective of this paper is to address systematic approaches to the study of certain unanswered questions in telemedicine. In particular, we will examine the problem from the perspective of health services research (HSR), which is "the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being" (Academy Health, 2004). HSR differs from clinical trials and small outcome studies, tending to focus primarily on the care delivery system itself, and not on the effects of specific interventions. It has considerable potential as a method of better understanding the effects and effectiveness of telemedicine (Bashshur, 1998).

Because the subject of telemedicine is very broad, including a large number of technologies and applications across all possible care delivery settings, a comprehensive discussion of HSR methods in telemedicine would far exceed the scope of this paper. For this reason, we have chosen to limit our discussion to services provided to individuals in their own residences—a subject of considerable interest among telemedicine stakeholders. We will focus on telehealth as an adjunct or substitute for home health visits, and on the use of telehealth technology for the management of chronic conditions in the home.

Taxonomy of Telemedicine Applications

Given the protean nature of telemedicine, its assessment requires careful specification not only of outcomes, but of the technology and clinical applications one is evaluating (Bashshur, 1995; Institute of Medicine, 1996). As noted above, there are many ways in which telehealth services might be provided. Each involves a different configuration of equipment, telecommunications media, and health care personnel. The patient may be present (synchronous) or not (asynchronous); a provider may be required on both ends of the interaction simultaneously, or the entire interaction might be handled by e-mail (asynchronous).

For this reason, Grigsby et al. (2004) developed a taxonomy of telemedicine applications for the Robert Wood Johnson Foundation, with the overall aim of permitting the rational classification of home telehealth services. More specifically, the taxonomy is intended to facilitate discussions of outcomes and costs in home telehealth, as well as to clarify the processes involved in care delivery, thereby permitting the implementation of fair and equitable reimbursement schemes. Although the taxonomy was intended to cover the various uses of telehealth technology in the home (or other residential setting), it could be expanded to include other non-residential applications (e.g., telerobotic laparoscopic surgery; Abbou et al., 2001; Mack, 2001). However, because our focus in the current paper is on telehealth in the home, we will discuss it briefly as it was presented originally.

The taxonomy categorizes the various health care services that can be provided in the home. The technologies are subdivided into synchronous (real-time interaction) and asynchronous (interactive, but not in real-time) categories. Synchronous technologies include IAV and interactive audio (telephone or radio), store-and-forward (SF) telehealth, e-mail, and data transmission. Asynchronous systems may mediate such processes as the transmission of blood glucose data via modem, or the in-home monitoring of activities of daily living (ADLs) and instrumental ADLs (IADLs; Glascock & Kutzik, 2000). The use of full-motion video may be asynchronous as well (e.g., systems that permit video monitoring of frail or cognitively impaired

individuals). The asynchronous technologies also include Web-based systems, through which patients and providers may exchange health data and/or other information. The reader should refer to the original publication for a graphical depiction of the taxonomy.

Frameworks for the Evaluation of Telemedicine

Bashshur (1995) was the first to propose a systematic approach to the evaluation of telemedicine. He argued that one might focus on "biomedical research, which encompasses issues of clinical effectiveness and safety" (p. 27), emphasizing specific technologies and clinical applications, and the assessment of their efficacy, effectiveness, reliability, and accuracy. Bashshur also highlighted the importance of HSR, which emphasizes such issues as cost, quality of care, acceptability, and access to care. From this second perspective, Bashshur recommended three consecutive stages of research in telemedicine: *evaluability assessment, formative evaluation,* and *summative evaluation.*

Evaluability assessment, which Bashshur argued "will frame the evaluation issues," setting "the stage for the systematic formative and summative evaluations to follow" (p. 27), is closely related to the measure of *structure* proposed by Donabedian (1980) and discussed below. Formative evaluation involves a descriptive study of system design and implementation, and measurement of short-term and intermediate-term effects on outcomes and on the way care is delivered. The emphasis of summative evaluation is on the measurement of end-result health outcomes of telemedicine systems (e.g., general health status, physiologic variables, functional status).

Grigsby and his associates (1995b) proposed a preliminary taxonomy of telemedicine applications that was based primarily on the processes involved in specific clinical applications rather than on disorders, specialties, or technologies (e.g., transmission of images, disease management, medical/surgical follow-up). Focusing on issues associated with the diagnostic accuracy of telemedicine, they recommended narrowing the scope of evaluation by using a relatively limited number of important or difficult to diagnose sentinel conditions, and suggested a set of five criteria for the selection of these conditions. Sensitivity and specificity analysis would be used to study accuracy, and standards for accuracy would be established as a function of the condition and its significance from the perspectives of personal and public health. This framework was embedded in questions that were particularly relevant at that time, especially regarding the diagnostic accuracy of teleradiology, which was then the most active area of research in telemedicine (Grigsby et al., 1995a). Although this approach remains useful, it is too circumscribed for application to many of the research questions facing the field today.

DeChant et al. (1996) argued in favor of a staged approach, grounded in the methods of technology assessment (e.g., Fineberg & Hiatt, 1979). Stage I involves the assessment of technical efficacy for specific applications and clinical endpoints (e.g., *is the system accurate and reliable in securely transmitting vital signs?*). Stage II involves the assessment of costs, quality, and access for specific applications and endpoints employing telemedicine systems that have been demonstrated to be technologically sound (in Stage I). Stage III represents a broader study of the effectiveness of *systems,* a more global evaluation that takes into consideration multiple endpoints and overall costs, in an effort to understand effects of telemedicine on the health care system.

A somewhat more detailed framework, also from the vantage point of technology assessment, was proposed by Ohinmaa et al. (2001; see also Ohinmaa & Hailey, 2002). Like other authors, they recommended beginning with technical assessment, followed by studies of effectiveness (e.g., diagnostic accuracy, changes in quality of life and in health), patient and

provider assessment of the technology (usefulness, usability, satisfaction), and costs. Although they suggested economic analyses as well, they noted the very real difficulties in conducting cost-effectiveness or cost-utility analyses of telemedicine at this time, given the rapidly changing technology, decreasing prices, and other factors (also Ohinmas et al., 2002; Grigsby, 1997). Ohinmaa and colleagues suggested that randomized controlled trials (RCTs) frequently may not be feasible in telemedicine, and that quasi-experimental methods may play an important role.

The most comprehensive evaluation framework to date was developed by a committee of the Institute of Medicine (IOM; Field, 1996), as a "base for strengthening individual evaluations of telemedicine and encouraging the coordination of evaluation strategies across projects and organizations, when possible." Toward that end, the IOM report proposed the integration of a number of elements into any telemedicine evaluation plan. Perhaps most notably, these include a business or project management plan that focuses on self-sustainability. This recommendation led to an emphasis on how telemedicine programs are to become self-supporting, and to a requirement by some federal funders (e.g., the Office for the Advancement of Telehealth, or OAT) of a business plan for all grantees. The IOM further recommended that evaluations of telemedicine compared with conventional care delivery target processes of care, health outcomes, access, costs from different perspectives (e.g., society, providers, payers), and perceptions of and satisfaction with telemedicine on the part of both patients and providers.

Overview of HSR

HSR is a broad and interdisciplinary field that examines the influence of the organization, delivery, and financing of health care services on accessibility, cost, and quality of care. Although research evaluating health services was underway as early as the turn of the 20th century, the term "health services research" was not coined until 1959. In the United States, the discipline began to develop rapidly after the passage of the Social Security Amendments of 1965. These amendments, which established the Medicare and Medicaid programs, dramatically increased federal expenditures on health care and thus the need for reliable information about the utilization, delivery, financing, and quality of health care services (Ginzberg, 1991). The expanded role of the federal government in the financing and provision of health care has had a direct effect on the nature of the field of HSR, which has grown dramatically and taken on a strong health policy orientation (Greenberg & Choi, 1983).

Accessibility of Care: In the context of health care, accessibility refers to the ease and timeliness with which patients are able to obtain and maintain needed medical services. Access to care is influenced by characteristics of the provider, the health care facility, and the patient (Donabedian, 1980). Major barriers to access include low socioeconomic status, underinsurance or lack of insurance, cost of care, and geographic location of patients, providers, and facilities.

Enhancing access to care has been one of the primary goals underlying the development of telemedicine programs. Early in its history, telemedicine focused heavily on making primary and specialist care available to individuals living in remote communities, where medical resources were limited in number and scope (Field, 1996). Although the field of telemedicine has broadened to focus on the provision of health care services even for patients for whom geographic isolation is not a concern, improving access to care continues to be an important goal for many telemedicine programs.

The accessibility of services has important implications for both the quality and cost of care. Greater access to health care is likely to lead to higher levels of service utilization and ultimately better health outcomes (presuming the services provided are medically appropriate), but may increase the cost of care as well (DeChant et al., 1996). In the context of telemedicine

evaluation, assessing the impact of a telehealth service on access, cost, and quality simultaneously provides a strong case for the overall impact of telecommunications and information technology on health care.

Measuring the accessibility of health care can be a challenge. In some cases, it is possible to evaluate accessibility directly, using objective measures of the ease and timeliness of care. For example, the delay in obtaining a dermatologic or radiology consultation may be far shorter when telemedicine services are used than when a patient from a rural location must travel to a remote health care facility for evaluation. Accessibility commonly is measured through the computation of less direct, proxy measures. For example, in an evaluation of a telemammography program, the proportion of eligible women who received screening mammograms as part of the telemedicine intervention could be compared with the proportion of women not in the telemedicine program who were appropriately screened. Although the percentage of women receiving mammograms is not a direct measure the ease or timeliness with which services are obtained, it is suggestive of appropriate accessibility to needed care.

Cost of Care: An assessment of the cost of care is an important component of telemedicine evaluation. Providers, funders, and policy-makers need access to information about the cost of telemedicine services relative to conventional health care, as well as information about the effect of telemedicine on access to and quality of care. There are numerous types of health care costs that are evaluated in telemedicine research, including direct costs, indirect costs, opportunity costs, intangible costs, and cost-effectiveness. In this paper, we will highlight direct medical costs related to the utilization of health care.

Health care utilization can be used as an indirect measure of both high and low quality health care. Utilization of services such as emergency room care and inpatient hospital care, can provide an estimate of the direct cost associated with poor medical or disease management. Conversely, utilization of other services, such as colorectal cancer screening, can reflect the direct costs associated with the appropriate use of preventative services.

Utilization of conventional health care services can provide important information about the quality of a telemedicine intervention. For example, researchers investigating the impact of a home telehealth disease management program for diabetes patients might compare their patient's utilization of emergency and inpatient services with the utilization of those services among a group of diabetics who have not received telemedicine services. If the telehealth intervention enhances the ability of patients to manage their diabetes, this effect may be apparent in a lower use of emergent and inpatient care.

One of the advantages of studying the direct cost of health care service use is the ready availability of utilization data. The federal government, as well as state governments and insurance companies, maintain databases of all claims filed, in their final, adjudicated form (Iezzoni, 1997). For example, the Home Health Agency Standard Analytic File maintained by the Centers for Medicare & Medicaid Services (CMS) includes information about each billable visit received by Medicare and Medicaid patients receiving home health services. These data could be used to determine whether patients involved in home-based telemedicine programs require fewer in-home visits than patients receiving only conventional care. Investigators would be able to examine the impact of the telemedicine intervention on specific types of home care services. For example, a particular telehealth intervention might be expected to reduce patients' needs for skilled nursing visits, while having little effect on the utilization of medical social work care. Similarly, the CMS-maintained Inpatient Standard Analytic File provides detailed information about the utilization of hospital services. This file can provide telemedicine

researchers with a means of examining the use of services that may suggest ineffective management of patient disease. The increased access to care that telemedicine programs may provide may result in reduced utilization events such as inpatient hospitalization.

Quality of Care: The study of health care quality has become an increasingly important focus in the field of HSR. Numerous definitions of quality have been proposed. Some definitions focus on positive health outcomes, such as the maximization of patient welfare (Donabedian, 1980) or the contribution of care to patient quality of life or longevity (American Medical Association, 1986). The IOM defines high quality care as health care that is compatible with current medical knowledge and that increases the probability that an individual will achieve a desired health outcome (Lohr, 1990). Other health services researchers propose a definition of quality that highlights the minimization of negative health outcomes, such as adverse events (Ellwood et al., 1973). Still others underscore the specific characteristics of medical care that are deemed to reflect quality (i.e., accessibility, availability, acceptability, appropriateness, continuity, competency, and safety; Roberts, 1988).

Avedis Donabedian (1980), one of the forefathers of research in health care quality, proposed three types of quality measures: structure, process, and outcomes. A structural evaluation of a given medical service provides an indirect measure of health care quality through the assessment of facility and personnel characteristics that are expected to influence quality. Researchers conducting structural evaluations focus on the adequacy of the facilities, equipment, providers, financing, and administrative processes within a particular health care setting. Evaluations of telemedicine services often have taken the form of structural evaluations, focusing on the adequacy and feasibility of telemedicine technology within a given health care setting.

Studies examining the process of health care are designed to assess the technical and interpersonal aspects of medical practice as well as patient behavior (Donabedian, 1980). To a large extent, process measures reflect the degree to which health care services are consistent with current medical knowledge and practice (i.e., has the provider employed diagnostic procedures and medical treatments that reflect the current state of the medical art, as in evidence-based guidelines?). Process measures also can be used to assess whether the provider's communications with the patient are consistent with normative expectations about patient-provider interaction. Patient factors that may influence the achievement of an ultimate health outcome, such as care seeking behavior, also can be considered process measures (Starfield, 1973).

A critical measure of health care quality is patient health outcome, which demonstrates the end-result impact of the services provided. The outcomes evaluated in a study can be short-term, of intermediate duration, or long-term. Short-term outcomes reflect the immediate impact of a health service. For example, a patient's blood glucose level following an insulin injection reflects the patient's short-term outcome in response to the treatment. Intermediate outcomes represent a longer time interval between the treatment and the outcome of interest and often are reflective of the process of care. These outcomes do not represent an ultimate health outcome, but instead reflect processes or behaviors that may enhance or detract from the patient's likelihood of achieving a desired health outcome. For example, a diabetic's compliance with a recommended dietary or exercise regimen could be considered an intermediate outcome of care. Long-term outcomes reflect the impact of health care services on patient health status after an extended period of time. Blindness, end-stage renal disease, and glucose-adjusted life years are examples of long-term outcomes for diabetic patients. Table 1 provides illustrative examples of structure, process, and outcome measures that might be valuable in the evaluation of telemedicine services.

Table 1: Illustrative Structure, Process, and Outcome Measures in Telemedicine

Structure	Process	Outcomes
• Speed and technical quality of electronic transmission • Adequacy of equipment • Skill of care providers in using technology • Cost of technology • Geographic accessibility of services	• Sensitivity of diagnosis (i.e., disease is correctly diagnosed) • Specificity of diagnosis (i.e., disease is correctly ruled out) • Treatment plan is consistent with evidence-based guidelines for care	Short-term • Blood glucose level Intermediate • Patient adherence with medical advice • Acceptability of services • Patient and provider satisfaction with care Long-term • Quality of life • Health or functional status

Deleted:

Although many researchers believe patient outcomes to be the ultimate measure of health care quality, evaluation of outcomes has limitations that make a focus on patient-level outcomes challenging and sometimes impractical (Donabedian, 1966; Starfield, 1973). The evaluation of the impact of a particular health service on long-term outcomes can require the collection of health status data over extended periods of time. To examine the impact of a home telehealth program for diabetes management, for example, a researcher might want to examine the impact of the intervention on long-term outcomes such as amputation or blindness. Such a study would require the staff and financial resources needed to track and retain study subjects and to collect outcome data for years into the future. Further, long-term outcomes are influenced by many factors, only one of which is the health care service under investigation. To demonstrate the independent effect of a telehealth intervention, other factors that may influence patients' long-term outcomes (e.g., patient socioeconomic status) must be measured and controlled in study analyses.

In comparison, short-term outcomes often are easy to measure and data concerning them cost relatively little to collect. However, the impact one would expect from a given health intervention may not be detectable through short-term outcomes. For example, use of short-term outcomes such as blood glucose level in the evaluation of the home telehealth diabetes management program would provide little evidence for the benefits of the program for controlling diabetes. Evidence of a positive impact on short-term patient outcomes is unlikely to convince health care providers, funders, and policy-makers of the value of the service provided.

In some cases, intermediate outcomes may be the most appropriate and feasible measures of patient outcomes. Although such measures require shorter data collection periods (and thus are less costly to collect) than long-term outcomes, intermediate outcome measures often are considered to have a level of validity far superior to short-term outcomes. A study showing increased patient compliance with a recommended dietary regimen as a result of a telehealth intervention may not demonstrate a direct link between the intervention and improved health status, but researchers, health care providers, and funders alike are likely to find the impact of the intervention on an intermediate outcome of known value to be compelling.

Although the ultimate goal of health care services is to improve or maintain patient health, evaluating services using patient outcomes often provides little information about the mechanisms underlying the effect of these services (Donabedian, 1966). If a telemedicine intervention with diabetic patients is shown to result in improved Hemoglobin A1c levels for patients receiving telemedicine services compared to a control group receiving conventional services, the researchers would have evidence for the beneficial effect of the program, but would not know what aspect of that program enhanced patient outcomes. Likewise, if the telemedicine intervention resulted in poor outcomes compared to the control group, the researchers would not know what aspect of the telemedicine program was deficient compared with conventional medical practice. Structure and process measures may be helpful in pinpointing the cause of the telehealth effect on patient outcomes.

Despite the challenges of evaluating the impact of health services on patient outcomes, outcome assessment remains a powerful and vivid measure of the quality of health care services. Ultimately, researchers must choose their outcomes variables carefully. In specifying the outcomes to be assessed in future telemedicine evaluations, researchers should identify clinically meaningful outcomes that are expected to be impacted by the specific intervention under investigation and that will be compelling to researchers, providers, and funders. The availability of valid and reliable outcome measures and the acceptable timeframe and cost for data collection naturally should inform these decisions. Studies addressing the impact of a telemedicine program on the three pivotal issues of accessibility, cost, and quality are especially critical in that they allow stakeholders to understand the full ramifications of a telehealth intervention. Policy makers and funders may be less likely to make ground-breaking decisions about (or might make different decisions about) the widespread implementation or coverage of telehealth services when cost, access, and quality data are presented separately than when they are analyzed together within a single study.

Methods for Evaluating the Effect of Telemedicine Services

Health services research employs a variety of methods to investigate the impact of health care services on accessibility, cost, and quality of care. This section describes three primary categories of research design that are commonly employed in the field (true experiments, quasi-experiments, and non-experimental designs) and provides illustrative examples of how each design might be applied to the evaluation of home telehealth.

True experiments provide the ultimate test of the impact of health care services. In a true experimental design (e.g., RCT),, each subject is randomly assigned either to a treatment group or a control group. Random assignment requires that each patient have an equal chance of being assigned to either of the experimental groups, thus assuring that the two groups are unlikely to differ in any systematic way at the outset of the study. Differences in the accessibility, cost, or quality of the services between the two groups, therefore, can be attributed to the experimental manipulation. Unlike all other research methods, a true experimental design allows for the demonstration of a causal relationship between the type of services provided and the outcomes of interest.

In the context of a telemedicine evaluation, a subject involved in a true experiment would be assigned to receive services through telemedicine or through conventional medical practice. For example, in an evaluation of a home-based diabetes management program, each new patient entering the program would be randomly assigned either to receive disease management services in traditional medical settings or to receive in-home equipment through which indicators of diabetic functioning could be transmitted for clinical review.

As one of the primary goals of many telemedicine programs is to enhance access to care, the measurement of the impact of the home telehealth program on the accessibility of care is important. Access to care might focus on the impact of the telehealth program on the appropriate utilization of regular diabetic screenings such as foot exams. Direct costs associated with appropriate, preventive exams as well as costs associated with negative utilization events could be examined. Outcome measures for a study of this nature might include short-term measures such as blood glucose level, intermediate measures such as patient satisfaction with care, and long-term measures such as health-related quality of life.

Despite the great power of true experimental designs to demonstrate the impact of a health care service, only a small number of telemedicine evaluations have met the rigorous standards of a true experiment (Hailey, Roine, & Ohinmaa, 2002). To demonstrate the effectiveness of telemedicine as an appropriate system of patient care, more studies employing this experimental method should be conducted. Research findings based on true experimental design will provide the strongest test of the impact of telemedicine services and will supply the most convincing evidence with which policy-makers and funders can evaluate the potential of telehealth services.

Although true experiments have the potential to provide a strong case for the value of telemedicine, studies employing true experimental methods require a rigor and control that can be impractical in the evaluation of some health programs. A particular challenge in the evaluation of telemedicine programs, in which interventions often are applied system-wide and sample sizes are small, is the requirement that subjects be randomly assigned to experimental groups. When random assignment is not possible, quasi-experimental designs can be used to examine the impact of a particular health care service. Like a true experiment, quasi-experimental designs involve the comparison of a treatment group and a control group. This method differs from true experimental design, however, in that subjects are not randomly assigned to those groups.

There are numerous ways in which the treatment and control groups in a quasi-experimental study may be developed. In a cohort or case-control study, the outcomes of a treatment group are compared with those of a control group. Unlike in a true experiment, the treatment and control groups reflect convenience samples of patients. For example, in a home telehealth evaluation, patients from one branch of a home care agency may be designated to receive telehealth services, whereas patients from a second branch may be designated to receive traditional home care services. Although this design allows for the comparison of patients receiving telehealth services with those who receive traditional medical care, the researchers cannot be confident that the difference in medical treatment caused any differences observed in eventual outcomes. It is possible that baseline differences between the two groups of patients, or between the two agency branches themselves, are the cause of any observed differences.

When a concurrent control group is not available (e.g., every home care patient in the agency of interest is receiving telehealth services), researchers can compare the outcomes of the treatment group with those of an historical control group (e.g., patients receiving care from the agency in the year prior to the implementation of a telehealth intervention). In this case, researchers must be capable of computing the same outcome measures for the telehealth patients and the historical control group. Acquiring parallel data for the treatment and control groups may require the time-consuming review of medical records, which often do not contain needed information. Alternately, administrative data can provide an excellent source of information about accessibility, cost, and outcomes for both current and prior patients.

Administrative data are especially comprehensive and accessible in the field of home health care. CMS requires that all Medicare-certified home care agencies submit Outcome and Assessment Information Set (OASIS) data on all patients at specified time points (e.g., start of care, 60-day recertification, discharge). The data set, which contains information about patient demographics, living situation, inpatient facility admission, and caregiver availability, as well as functional, physiologic, and cognitive/behavioral status, is a rich source of information with which long-term patient outcomes can be computed and risk-adjusted. As mentioned previously, the Home Health Standard Analytic File maintained by CMS provides detailed information about billable home health visits, allowing for the computation of measures of service use, including total number of visits and number of visits by type (i.e., skilled nursing).

In absence of an appropriate historical control group, study patients may be used as their own controls using a pretest-posttest design. Using this quasi-experimental method, patients would be assigned to receive conventional services for a period of time, followed by telemedicine services for an equivalent time interval. This approach is beneficial in that the control group and the treatment group are identical (or nearly so) at the beginning of data collection during each time period. Because the same subjects are in each group, it is unlikely that the two experimental groups will display systematic differences in health that could influence the study results. The disadvantage of such a pretest-posttest design is that each phase of the study must be of fairly brief duration. Therefore, investigators using this approach are unlikely to be able to examine the long-term effects of telehealth on access, cost, and quality of care and must instead focus on outcomes of a more circumscribed nature.

Non-experimental designs differ from true experiments and quasi-experimental designs in that they lack a control group. Common non-experimental methods are the case study, case series, and correlational design. In a case study, the medical experience of a single patient is described in detail. A case series provides a similar synopsis of several consecutive patients. Both methods are observational and descriptive, involving no manipulation of variables or comparison to a control group. A correlational study involves the examination of the relationships between variables for a larger group of subjects. For example, one might examine the relationship between patient satisfaction with telemedicine services and compliance with medical recommendations provided by remote care providers.

Although non-experimental studies are incapable of demonstrating the causal effect of a given health care service, and thus are not perceived to be as credible and persuasive as their more rigorous experimental counterparts, they can provide important descriptive information about the services under investigation. These techniques are particularly useful when examining a new type of health care service or a service for which research evidence is scant. Outside of these situations, however, well-controlled experimental designs are far superior and should be the goal of researchers investigating the impact of telemedicine services. Policy-makers, providers, and funders will require stronger evidence than non-experimental studies can provide about the beneficial effects of telemedicine before services will become more widely accepted and available.

Means of Addressing Small Sample Sizes: Telemedicine evaluations often are limited as a result of their small sample sizes. Meta-analysis and collaborative research groups have the potential to address this limitation and provide telemedicine stakeholders with stronger evidence of the impact of telemedicine on access, cost, and quality of care. Meta-analysis is a quantitative method of combining the results of multiple independent investigations of the same phenomenon. A study of this sort involves the pooling and subsequent statistical analysis of the

results from multiple studies, allowing more definitive conclusions to be drawn than might be possible from each individual study examined. This method is particularly helpful when the studies that have been conducted previously involve small sample sizes, research designs that are methodologically limited, and/or that report inconsistent results. For example, Lee et al. (2003) used meta-analytic techniques to examine the relationship between physical activity and the incidence of stroke. Although the existing literature in this area provides inconsistent findings regarding this relationship, the meta-analysis conducted by Lee and colleagues showed a clear beneficial effect of physical activity on stroke risk. The use of meta-analysis in telemedicine evaluation has the potential to dramatically enhance the strength and persuasiveness of the empirical evidence supporting the use of telemedicine as an appropriate mechanism of providing patient care.

The development of research cooperative groups also has the potential to improve sample sizes. A cooperative group is a large network of institutions and researchers that jointly collect data following the same research protocols. The involvement of multiple data collection sites allows for the collection of a larger sample of data for a given study than each individual researcher or research institution would be able to collect independently. Larger samples sizes serve to enhance the statistical power of a study, as well as the accuracy and persuasiveness of the results. Cooperative groups have experienced great success in fields such as oncology, in which the Eastern Cooperative Oncology Group (ECOG) and Southwest Oncology Group (SWOG) have established networks of thousands of cancer researchers and clinicians across hundreds of institutions. ECOG alone enrolls approximately 6,000 patients each year in approximately 90 clinical trials.[1]

Collaborative research in telemedicine would allow for the development of larger datasets on which to base conclusions about the impact of telehealth services, and would encourage the testing of telemedicine interventions across multiple settings, thus providing more convincing and generalizable results than a study conducted in a single health care setting. Although collaborative research groups have the potential to be of great use in telemedicine, little collaborative work has been done. In the 1990s, two efforts were made to establish telemedicine collaborative research groups. The Telemedicine Research Center in Portland, Oregon organized the Clinical Telemedicine Cooperative Group (CTCG) with the goal of encouraging the development of multi-site telemedicine evaluations, the use of standardized data collection instruments, and the pooling of data across studies (Perednia, 1995, 1996). In 1993, the National Consortium for Telemedicine Evaluation was founded by Rashid Bashshur at the University of Michigan with the goal of conducting multi-state evaluations of telemedicine's impact on access to and quality of care. Although the foundation for collaborative research was laid during this time, neither organization was able to implement large-scale telemedicine evaluations as had been hoped. Given the need for strong evidence of the effects of telemedicine on access, cost, and quality of care, renewed efforts to establish cooperative research ventures, particularly those focused on the use of true experimental design, are warranted.

Discussion

The methods of health services research provide a potentially valuable framework for the assessment of telemedicine. The level of analysis generally is broader than is the case for clinical trials, with a focus on the system of care, and on outcomes, costs, access, and similar dependent

---

[1] This information is from the ECOG Website, *Introduction to ECOG,* (http://ecog.dfci.harvard.edu/general/intro.html). Information about SWOG is from the SWOG Website (http://swog.org/visitors/whatwedo.asp).

variables. Structure, process, and outcomes are typical domains of measurement, and the first two classes of variables may be used as both dependent and independent variables. For example, adherence to diabetes management recommendations, a short- or intermediate-term process outcome, also mediates longer term outcomes such as reduced incidence of diabetic retinopathy or end-stage renal disease. The conduct of HSR studies of many areas of telemedicine is difficult given the relatively low patient volumes available.

We focused on home telehealth in this paper in part to narrow the scope of the discussion, but also because this is a clinical application that is particularly well-suited to the methods of HSR. Administrative data are available from CMS regarding utilization (home health claims data), structure variables (home health agency administrative and cost data), and outcomes (the OASIS). HSR is unfamiliar to many investigators who conduct clinical research, but the field's approach, sources of data, and methods are capable of yielding valuable data that may undergird policy recommendations.

References

Abbou C, Hoznek A, Salomon L, Olsson LE, Lobontiu A, Saint F, Cicco A, Antiphon P, Chopin D: (2001) Laparoscopic radical prostatectomy with a remote controlled robot. *The Journal of Urology,* 165, 1964-1966.

Academy Health: (2004) *Glossary of terms commonly used in health care.* http://www.academyhealth.org/publications/glossary.pdf. Accessed 28 April 2004.

American Medical Association Council of Medicare Service: (1986) Quality of care. *Journal of the American Medical Association,* 256, 1032-1034.

Bashshur RL: (1995) On the definition and evaluation of telemedicine. *Telemedicine Journal,* 1, 19-30.

Bashshur RL: (1998) Rethinking the evaluation and priorities in telemedicine. *Telemedicine Journal,* 4, 1-4.

Bashshur RL: (2001) Where we are in telemedicine/telehealth, and where we go from here. *Telemedicine Journal and e-Health,* 7, 273-277.

Benschoter R: (1967) Multipurpose television. *Annals of the New York Academy of Sciences,* 142, 471-478.

Conrath DW, Dunn EV, Bloor WG, Tranquada B: (1977) A clinical evaluation of four alternative telemedicine systems. *Behavioral Science,* 22, 12-21.

DeChant HK, Tohme WG, Mun SK, HayesWS, Schulman KA: (1996) Health systems evaluation of telemedicine: A staged approach. *Telemedicine Journal,* 2, 303-312.

Donabedian A: (1966) Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly,* 44, 166-203.

Donabedian A: (1980) *Explorations in Quality Assessment and Monitoring. Volume I: The Definition of Quality and Approaches to its Assessment.* Ann Arbor, MI: Health Administration Press.

Ellwood PM, O'Donoghue P, McClure W, Holley R, Carlson RJ, Hoagberg E: (1973) *Assuring the Quality of Health Care.* Minneapolis, Minnesota: Interstudy.

Field MJ (Ed.): (1996) *Telemedicine: A guide to assessing telecommunications in health care.* Washington, D.C.: National Academy Press.

Fineberg HV, Hiatt HH: (1979) Evaluation of medical practices: The case for technology assessment. *New England Journal of Medicine,* 301, 1086-1091.

Ginzberg E: (1991) *Health Services Research: Key to Health Policy.* Cambridge, MA: Harvard University Press.

Greenberg JN, Choi T: (1983) The role of the social sciences in health services research: An overview. In T. Choi T & J.N. Greenberg (Eds.), *Social Science Approaches to Health Services Research.* Ann Arbor, Michigan: Health Administration Press, 3-20.

Grigsby J: (1997) Cost-effectiveness analysis and telemedicine: What do we know? In *Telemedicine Sourcebook,* New York: Faulkner & Gray.

Grigsby J, Kaehny MM, Sandberg EJ, Schlenker RE, Shaughnessy PW: (1995a) Effects and

effectiveness of telemedicine. *Health Care Financing Review,* 17, 115-131.

Grigsby J, Rigby M, Hiemstra A, House M, Olsson S, Whitten P: (2002) Telemedicine/ Telehealth: An international perspective. The diffusion of telemedicine. *Telemedicine Journal and e-Health,* 8, 79-94.

Grigsby J, Schlenker RE, Kaehny MM, Shaughnessy PW, Sandberg EJ: (1995b) Analytic framework for evaluation of telemedicine. *Telemedicine Journal,* 1, 31-39.

Hailey D, Roine R, Ohinmaa A: (2002) Systematic review of evidence for the benefits of telemedicine. *Journal of Telemedicine and Telecare,* 8 (Suppl. 1), 1-7.

Iezzoni LI: (1997) Assessing quality using administrative data. *Annals of Internal Medicine,* 127, 666-674.

Lee CD, Folsom AR, Blair SN: (2003) Physical activity and stroke risk: A meta-analysis. *Stroke,* 34, 2475-2482.

Mack MJ: (2001) Minimally invasive and robotic surgery. *Journal of the American Medical Association,* 285, 568-572.

Lohr KN: (1990) *Medicare: A Strategy for Quality Assurance.* Washington, DC: National Academy Press.

Ohinmaa A, Hailey D: (2002) Telemedicine, outcomes and policy decisions. *Disease Management and Health Outcomes,* 10, 269-276.

Ohinmaa A, Hailey D, Roine R: (2001) Elements for assessment of telemedicine applications. *International Journal of Technology Assessment in Health Care,* 17, 190-202.

Ohinmaa A, Vuolio S, Haukipuro K, Winblad K: (2002) A cost-minimization analysis of orthopaedic consultations using videoconferencing in comparison with conventional consulting. *Journal of Telemedicine and Telecare,* 8, 283-289.

Perednia DA: (1995) Telemedicine system evaluation and a collaborative model for multicentered research. *Journal of Medical Systems,* 19, 287-294.

Perednia DA: (1996) Telemedicine system evaluation, transaction models, and multicentered research. *Journal of AHIMA,* 67, 60-63.

Roberts JS: (1988) Quality health care: Its definition and evaluation. In E.F.X. Hughes (Ed.), *Perspectives on Quality in American Health Care.* Washington, DC: McGraw-Hill.

Rogers EM: (1995) *Diffusion of innovations.* 4th ed. New York: The Free Press.

Starfield B: (1973) Health services research: a working model. *The New England Journal of Medicine,* 289, 132-136.

Evolving Telemedicine/eHealth Technology - 21st Century

Frank E. Ferrante, MSEE, MSEPP
President/Founder FEF Group, LLC
Past Chair, IEEE-USA Medical Technology Policy Committee
Chair, Geriatric Care IT WG
Associate Faculty, JHU SPSBE
Executive Partner, The College of William & Mary

20 May 2004

Abstract

This paper addresses emerging technologies being applied to support a rapidly changing and expanding scope of telemedicine applications in this 21st century environment. The technologies of primary focus in this paper include wireless and its emerging broadband offering, nanotechnology, Intelligent Agent applications, and Grid Computing. Changes underway in wireless designs aimed at enhancing security are described. Some of the current work underway in developing nanotechnology applications, research into the use of intelligent agents to establish what are termed Knowbots are described, and a sampling of the use of Grid computing capabilities in support of medical applications are addressed. Finally, a discussion is provided on the expansion of the applications of telemedicine that reflects a response to the need for cost containment in healthcare and to support ongoing changes seen in our healthcare demographics towards a more aging and mobile population.

Introduction

As the global Internet evolves into today's more highly mobile and broadband service offerings, it is anticipated that the applications of these new services to support telemedicine and eHealth operations will result in increasing healthcare benefits for all and at a lower cost. This paper, representing an extension of our earlier documentation of telemedicine technology offerings in 2001 time frame[1] , describes state-of-the-art technologies that are expanding telemedicine's offerings world-wide.

The following sections provide an overview of the more recent wireless characteristics and capabilities that need to be considered when reviewing its application as a segment of an end-to-end telemedicine service arrangement. It is noted that wireless services generally represent only a segment of the end-to-end linkage needed to support a telemedicine network. However, the changes that have been occurring in the wireless services area appear to be the more dynamic of all segments leading to potentially the greatest savings in some instances given that the service's reliability, security, and quality as well as data handling performance meets the rigorous requirements of our medical information handling configurations. For this reason, considerable

attention is being focused on this technology over that of the wireline offerings. Fiber optics, cable, twisted pair wiring, long-haul telecommunications services, and comparable switched and dedicated offerings from our national and international telecommunications carriers, as well as satellite service offerings all continue to require attention when designing end-to-end telemedicine networks.

In addition to wireless, two other relatively new technologies are entering our sphere of visibility for their possible consideration in designing telemedicine networks. One, more at a molecular level of design, is nanotechnology. Nanotechnology is an emerging interdisciplinary area of science that focuses on the study of structures of nanoscopic dimensions and their scientific and technical properties.[2]

Within the scope of nanotechnology there is a synergy between micro techniques, medicine, and biology that when applied properly result in new products at the scale of a single cell. These new products, one class of which is called MEMS (Micro-Electro-Mechanical Systems), are extremely miniaturized. Because they replace some more cumbersome or externally applied biomedical sensors and diagnostic tools, MEMS products are being shown to be of key importance for improved diagnosis and therapy both now and in the future. MEMS will be discussed in the sections that follow.

Finally, we describe the application of Intelligent agents in forming Knowbots that can be used in cognitive situations and we discuss future approaches to better utilization of our global computer processing capabilities for medical analysis, Grid Computing.

Technology Overview

Table 1 illustrates the relative image sizes of uncompressed digital representations of medical images that are transferred as part of our telemedicine network. Correspondingly, Figure 1 presents a sample of transmission times required when highly data intensive mammography images (sized at approximately 24 Megabytes (MB) - uncompressed - A Byte is 8 binary bits) are transmitted over links offering varying access speeds. If a typical voice line at 64 kilobits per second (kbps) were to be used for the image transfer, Figure 1 indicates that it would take slightly less than an hour per image for transmission (50 minutes). However, given today's 11 Megabits per second (Mbps) wireless WiFi (Wireless Fidelity) capabilities the transfer time of a mammogram can be reduced to less than a minute per image. This illustrates the higher capability now possible using improved wireless offerings that are available.

Table 1 – Digital Image Sizing Estimates*

Image Type	Image resolution		Image Size
	Spatial	Size(bits/pixel)	less Control & error bits
Ultrasound	512x512	x8	256 Kbytes
Other (Angiography, Endoscopy, Nuclear Med., Cardiology, Radiology)	512x512	x8	256 Kbytes
Computed Tomography	512x512	x12	384 Kbytes
Magnetic Resonance Imaging	1024x1024	x12	1.5 Mbytes
Digitized (Scanned) X-Ray	1024x1280	x12	1.9 Mbytes
Digital Radiology	2048x2048	x8	4 Mbytes
" " (high quality)	2048x2048	x12	6 Mbytes
Mammography	4096x4096	x12	25 Mbytes

*Revised from Figure 3 of reference 1 (Byte equals 8 bits)

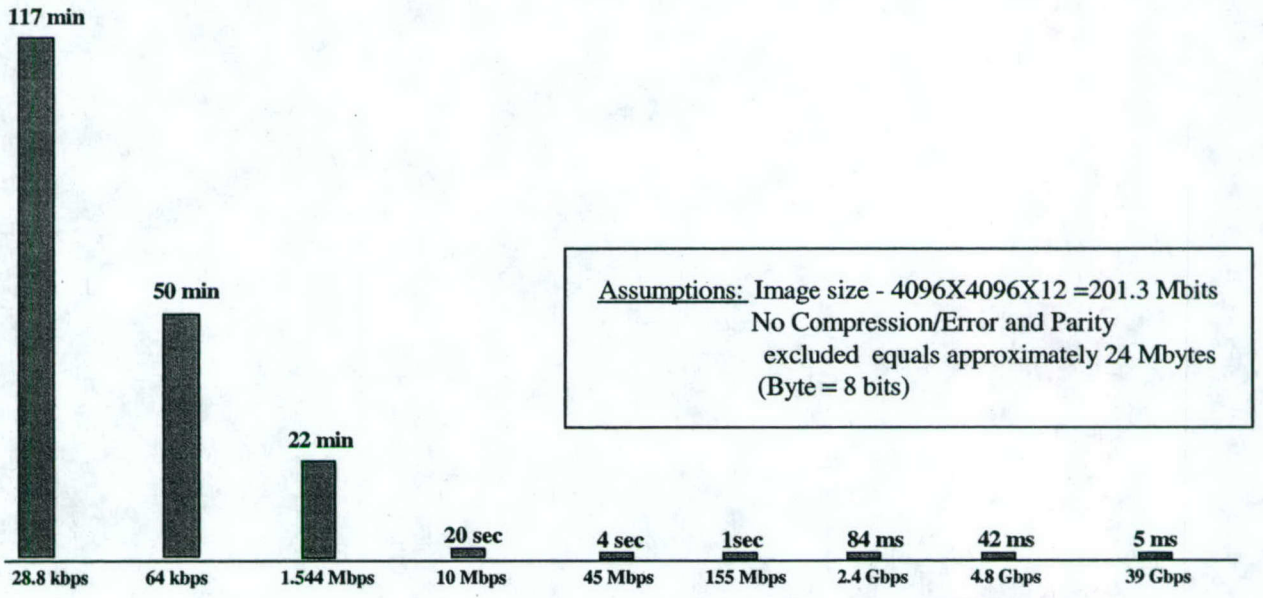


Figure 1- Image Transfer Times (in minutes, seconds, and milliseconds)

As an even better choice, data transfer rate options, now being offered from the latest Institute of Electrical and Electronics Engineers (IEEE) standard under review for WiFi systems, are as high as 54 Mbps (see "optional" data rates offered in reference 4). Choosing this higher optional rate (54 Mbps) over the 11 Mbps reduces the image transfer times even more - from slightly less than an hour to slightly less than 4 seconds per mammogram – making this an even better option for future telemedicine systems.

Cost of implementing these higher quality wireless arrangements in a local area environment can be as little as a few hundred dollars for system equipment making these very reasonable for small facility operations

Looking at this offering from an economic perspective, the cost per bit of services interconnecting the local area environment at these higher data rates has also benefited the user through competition – cost has dropped considerably. In 1996 the cost for a T1 (1.544 Mbps) service arrangement between Washington, DC and Los Angeles, CA was determined to average between $0.0025 per Mbit to as low as $0.0005 cents per Mbit[3]. Today's dedicated T1 comparable service has fallen to approximately half this monthly cost. Non-dedicated Internet service provided by Internet Service Providers for residential as well as business subscribers (e.g., DSL, cable, and VSAT satellite) can be procured for less than $60 per month, thus reflecting a per Mbit rate of approximately $0.00003* (a reduction by a factor of ~17). The following sections describe the characteristics of this dramatic change in technology and cost and review the drivers anticipated to speed up the move to this .

Wireless

One simply has to observe the omnipresent cell phone to notice that something extremely vital to our daily lives is changing. That something is mobility. The Internet of the future is now moving ever so rapidly from a fixed end-to-end operation of the 80's and 90's to a more mobile design of the 21st century variety. Cell phones usage is not telemedicine's solution however. Cell phone expanded usage merely reflect a phenomena that points us to where change is occurring – we now must determine how to apply this to our highly critical needs in a cost effective and beneficial manner.

Generally, it is not the cell phone that is our focus, it is the technology of having ubiquitous wireless (untethered) access to the global interconnecting Internet services in a manner that, if proven to be cost beneficial and secure (meaning that it offers us the required ability to maintain the privacy of information being processed through this media), then this is the wave of the future for telemedicine to include in its designs.

Table 2 lists the latest standards recommendations that the Institute of Electrical and Electronics Engineers (IEEE) have issued or are in the process of issuing (e.g., 802.11i) and which are already being included by manufacturers in their wireless product designs. The 802.11a, b, and g are commonly referred to as the WiFi recommendations.

---

* **Assumptions:** $60 per month charge for Internet DSL/Cable/Satellite service average; ~4 Million seconds per month full time usage vs. $2000 per month fro a dedicated T1 line. (DSL – Digital Subscriber Line; T1 – 1.544Mbps)

Table 2- Wireless IEEE Standards Recommendations subgroups[4, 5]

	802.11 Standard Subgroups - Primary interest			
802.11a	High Speed Physical Layer in 5GHz Band			
802.11b	Higher Speed Physical Layer Extension of Wireless in 2.4 GHz Band			
802.11d	Local and Metropolitan Area Wireless			
802.11g	Broadband Wireless			
802.11i	802.11 security			

*GHz – Giga-Hertz (Cycles per second)

These three 802.11 subgroups offer a wide range of wireless services that can be applied in a telemedicine environment both in a home setting as well as office. Figure 2 depicts the primary characteristics of the 802.11 WiFi standard. Note the additional data handling rates now included within the 802.11b and g portions of the wireless standard. Due in part to the extensive popularity of this in-home wireless WiFi product line using 802.11b, a Working Group was established to extend the number of mandatory data rates being offered by the products to meet higher demands for faster information transfer rates over the wireless systems. As a consequence, 802.11b, which previously was limited to offering only two choices in data transfer rates (5.5 and 11 Mbps), was extended to allow at least 8 more choices of data transfer rates (802.11g offerings). From a telemedicine perspective this is extremely useful in that many of the higher data rates would allow for speedier transfers of the large imaging files mentioned earlier. In addition, because of the unique wireless protocols applied (referred to as the Spread Spectrum techniques in Figure 2) some added security as well as interference protection for wireless transmissions were found to be inherent in the signaling applied. This is of value to the medical service in that it offered a level of higher protection for the information being transmitted and helped in assuring that the signals were more readily acquired or less apt to be interfered with by other systems operating within a medical facility.

Figure 2 – 802.11 WiFi Standard Characteristics[4, 5]

IEEE 802.11			IEEE 802.11a	IEEE 802.11b, [g]
2.4 GHz Frequency Hopping Spread Spectrum 1 Mbps 2 Mbps	2.4 GHz Direct Sequence Spread Spectrum 1 Mbps 2 Mbps	Infrared 1 Mbps 2 Mbps	5GHz Orthogonol FDM 6,9,12,18,24, 36,48,54 Mbps	2.4 GHz Direct Sequence Spread Spectrum 5.5, 11 Mbps [+1, 2, 6, 12, 24 Mbps Mandatory; 36, 48, 54 Mbps Optional - under 802.11g]

Another possible offering that is considered when working in a hospital environment, where the choices of frequencies being used by a wireless service may be critical due to potential interference, is that offered under 802.11a. This part of the 802.11standard offered even more potential protection from interference for medical applications since it's operational frequency was within a much higher band than many of the other medical systems with the medical environment (5 GHz). Many consider 802.11a operations as the industry preferred application for wireless.

Wireless security

In considering the application of wireless networks for telemedicine and the need for protecting patient medical information, "Security" is considered a top priority. In the United States, current HIPAA (Health Insurance Portability Assurance Act) regulations require that precautions be taken to ensure that patient medical information is available to those who need it, and that it is protected against those lacking proper credentials, and not modified — either intentionally or unintentionally[6].

As Davie remarked in his referenced article[7]: "...Security is one area where health care places stronger demands on technology than almost any other application area because of the severe and irreversible consequences of improper disclosure or modification of information. " For this reason and for the many others that Davie remarked on in his article, the following discussion on the merits of the 802.11 Wired Equivalent Protection (WEP) security offering and its improved successor (802.11i offering) that is now being offered by many in their hardware designs, is a necessity to understand by those designing telemedicine systems.

From a security perspective, none of the 802.11 standard products that were "initially" developed could be considered as secure due to potential weaknesses in the WEP security offerings that were associated with 802.11.

The IEEE 802.11 wireless LAN standard defines an authentication and encryption service based on WEP. Note the term "wired." This implies that the search for a security schema was focused on providing the over-the-air wireless service the same protection that might have been afforded to a basic wired telecommunications service (note, wired services can be controlled and monitored for intrusion whereas wireless system cannot.)

To understand the weaknesses in WEP we should review some of the relatively basic characteristics of the process employed. WEP generally offers use of several secret key authentication levels. This authentication "secret" key when used in conjunction with a "public" key and as part of the Public Key Infrastructure (PKI) encryption process supports not only encryption of digital traffic, but it also serves to support the other main attributes of a secure transmission system, namely,

- Confidentiality – ensures that message sent can only be read by authorized parties
- Authentication-Ensures that origin of message is correctly identified
- Integrity- ensures that only authorized parties are able to modify information
- Non-repudiation –Neither the sender nor the receiver can deny receipt or sending of the

transmission

- Access Control – requires that access to information may be controlled by or for a target system

All of these are critically important in the use of a wireless operation within a medical records environment since the current HIPAA regulations require the maintaining of patient information as private and security means provided by a medical professional are one of the means for maintaining privacy if medical information is transmitted over the wireless airways.

For emphasis, it is repeated here that the WEP process, though acceptable for many home and small business systems, is insufficient to secure a telemedicine system design requiring full privacy of patient healthcare and personal medical records. Several reasons for this are known. The process of passing encrypted information over open airways seems at first glance to be very secure. For a personal local environment, this may be considered adequate. However, if your system is one that is being offered to support private medical record information, tighter security means must be established.

To understand the weakness within WEP and thus enhance the service with a more secure offering one must review some of the WEP processes, at least at a fundamental level of complexity. WEP uses an Initialization Vector (IV) - a 24 bit pseudo-randomly generated code - that serves as a seed for developing its encryption process. This IV unfortunately employs a card that when it is first inserted into a computer, is set to zero and subsequently is simply incremented by 1. Some vendor products that introduce the IVs for its offering operate off of a pseudo-random number generator package built-into the computer system provided rather than initialize itself from 0. This was intended to offer a semblance of improved security. However, even when this change in added, there remains the fact that there are only 24 bits to an IV and this means that there is a probability that reuse of the IV every so often must occur (usually within 5000 cycles). As a result, a determination of at least one of the IVs by a potential intruder can be made within minutes of operation leading to an initial stage in a process of breaking all codes used by the system[8]. Several programs exist for performing this action. As a second level of security failure, the WEP process uses what was once thought to be an excellent cipher encryption algorithm to perform its encryption referred to as the RC4 stream cipher algorithm. Weaknesses in this algorithm were identified as far back as 2001[9] and since then, these weaknesses allowed the smart intruder to break the code and gain information almost immediately.

To offset some of the weaknesses found within the current WEP algorithm, designers have attempted to introduce several additional enhancements to security when using wireless. One approach is simply to specify that only equipment identified by specific Media Access Control (MAC) addresses (each computer or hardware element in a network has a unique MAC address that it is assigned) are allowed on the wireless links. Without this address being entered explicitly by the administrator of the system into a list that authorizes wireless access, the wireless system becomes effectively undetectable. This approach is reasonable for some home users and possibly small business users, but again, it is insufficient for assuring security of a medical record support system since methods of spoofing MAC addresses do exist.

Finally, the IEEE established a working group to address the security issues identified as weaknesses within its 802.11 standard. This activity, referred to as the 802.11i Working Group, has itemized the weaknesses identified and developed a means to counter each of the items securely and in a standard manner that would allow all manufacturers to interoperate effectively. Some of the typical vulnerabilities[10] that have been recognized by the IEEE within 802.11 include:

- No per-packet authentication
- Vulnerability to disassociation attacks
- No user identification and authentication
- No central authentication, authorization, and accounting support
- Stream cipher scheme (RC4) is vulnerable to known plain text attacks
- Some implementations derive WEP keys from passwords
- No support for extended authentication; for example: token cards; certificates/smart-cards; one-time passwords; biometrics; etc.
- There are key management issues; for example, re-keying global keys, and no dynamic, per-session key management

This standard, still in recommendation stages of development, is currently being included in many of the common product lines now available. It is highly recommended that this standard be employed as the guideline for future telemedicine applications where wireless technology is employed.

Nanotechnology in Medicine

Nanotechnology is envisioned as the pill of the future - - An individual swallowing a pill sized product and from a remote location a medical professional monitors and tracks the pill as it wanders through the body's systems in an organized manner, looking for polyps, scanning for cancer cells, checking blood cells and infectious conditions in the blood stream, or more directly providing a video view of the inside of a patient's arteries and veins to confirm that a blockage is about to happen. If the pill uncovers something suspicious to the physician, it can be stopped and through an external RF signal, it is directed to re-examine or even medicate the spot of interest. The pill may save the patient from painful situation by allowing visual view by a physician without requiring invasive surgery. This process is the next generation in telemedicine. Providing devices of the 21st century to offer patients an alternative to currently used invasive techniques. Sounds far fetched! Well, maybe not.

Nanotechnology is a multidisciplinary scientific technology that applies to products or devices that are manufactured at a molecular level[11]. From a medical perspective, nanotechnology represents a synergy between micro techniques, medicine, and biology that can result in new products at the scale of a single cell. New products, one class of which is called MEMS (Micro-Electro-Mechanical Systems), is extremely miniaturized and because they supplant some more cumbersome or externally applied biomedical sensors and diagnostic tools, they are being found to be of key importance for improved diagnosis and therapy both now and in the future.

From our previous paper on Telemedicine Technology[1], we learned that MEMS devices are but one currently existing set of devices that can be included within the definition of a nanotechnology application representing new opportunities for telemedicine. MEMs were noted to take many forms from small elements of a robotic device used to assist in non-invasive surgery, to an encapsulated camera. Wearable wireless sensors the size of a spec of dust were also listed under the category of a potential nanotechnology application. Memory glasses, a wearable, context aware, proactive memory aid offers something similar to the projection of the device in today's real world.[12] This device, described along with several others is in prototype form at this time. It serves to provide messages to the wearer to take medication at certain times, do exercise, or even remind the wearer of the names and information of friends that are separately listed in a nearby computer file (connectivity to the computer could be via an 802.11 wireless connection, by means of a simple infrared link between two interoperable devices, or through a third wireless source called Bluetooth).[13, 14]

In a recent article by Alex Pentland of MIT, he mentions some of the newer devices now under development. These include:

- nano-particles - that can be used for diagnostic and medical screening purposes
- artificial receptors
- DNA sequencing devices using nano-pores
- uniquely developed drug delivery systems that can be injected long-term and programmed to provide medication as prescribed without physically having to digest pills every day
- gene therapy applications
- tissue engineering enabling devices using cell size nano-devices formed at the atomic level.[15]

In addition to this relatively small set of devices, Freitas has listed and described almost a dozen more devices in nanomedicine that are representative of the future.[16] Freitas descriptions include:

- Nanotweezers controlled by voltages and potentially capable of retrieving items inside a person that would normally take surgery to acquire (similar to our conjectured view of nanotechnology application sin our opening paragraph of this section).
- Nanometers
- Nano-computers
- Nano-robots capable of supporting chromosome replacement therapy
- Nanopore Sensors with DNA sequencing that can be externally regulated by a voltage or RF controlling device

These devices and processes lead the way to the next generation of what we will be using in our advanced telemedicine environment.[16]

The start of this nanotechnology appears to go back as far as 1959. Emerich[11] describes a meeting of the American Physical Society in which Dr. Richard Feynman lectured on designing machines that were programmed to build smaller machines in a process beyond what man could do

manually.[15] This was a vision of what we now refer to as nanotechnology.

The development of miniaturized computers on a chip are passed off as one form of nanotechnology. One robotic-like device developed recently by Intel is a computerized walking cane developed recently by Intel that monitors the pace and cantor of a patient and remotely records the steps and pressure used to help in making an intelligent determination as to when a fall may occur.[17]

In a recent paper by Dishman he describes several breakthrough nanotechnology related devices and systems currently in place.[18] From a biological and sensor level, mobile, embedded, wearable and even implantable technologies are used to measure sleep patterns, monitor eating habits, measure body temperature and blood pressure and diagnose all of these to provide a signature for diseases. Dishman describes proactive computing systems that monitor behavior and offer advice to the patient or physician that it predicts will keep someone like the aging from becoming inactive or making improper decisions.

Cognitive issues with the aged are also using nanotech devices aimed at providing guidance to those needing it so that they do not have to be moved into a costly nursing facility too soon. The longer one stays at home, the less the cost to both the individual as well as to the public. Costs for Assisted living today vary from as low as $10,000 per year in subsidized charitable facilities to the average of $25,000 per year and higher (over $100,000 is many major markets given the service needed to assist Alzheimer patients. A number of robotic devices have been developed and researched by Dr. Martha Pollack from University of Michigan. Pollack's devices work to compensate for physical and sensory deficiencies occurring in an individual entering some of the early stages of dementia[19]. Her work is addressing the cognitive processes directly.

Grid Computing in Medicine

What is Grid Computing? Answer: A family of technologies for dynamically and opportunistically provisioning computing power from a pool of resources.[20]

This technology is considered critical to our education in advanced technologies in that it may allow us to perform analysis not limited in speed and resources to our own meager computer system or host offerings. Grid computing, basically allows us access to a grid of interconnecting systems that permits us to increase our processing capabilities by multiple orders of magnitude. This potentially offers us a means to reduce time for analysis, which in turn may open up new avenues of study for the next generation of telemedicine.

Grid computing allows us to take advantage of the millions and possibly billions of systems out in the global environment that have time to offer their processing capabilities for us to use without impacting personal use of an individual system. A grid of computer system, operating in a secure manner, would allow us to all have super computing capabilities from our own local environment. It can be the mechanism to accomplish the highly computationally intense study of cognitive processing heretofore not envisioned except through a singly massive super computing system. These systems are closely held by large Universities or companies due to their cost and management requirements. Now, through grid computing, the cost may be shared by millions

where their individual computer has spare cycles now used in a secure, and private manner that would be safe from intrusion by new layers of software and hardware built into each of the grid's new computer.

The question now being asked is how real is grid computing? Is this happening now? Is it possible to make one computer work like this with another, with thousands of others in a secure and non-intrusive manner?

Investigating these questions, the following can be said. Yes, Grid Computing is real, it is happening now and secure interconnections are possible given the right steps are followed. [20, 21]

Scientists and engineers have used computing grids for more than two decades. Millions of desktop PCs run a grid application behind a popular screen saver called SETI@home. screen saver is but one example of grid computing. "...SETI, the Search for Extraterrestrial Intelligence (http://setiathome.ssl.berkeley.edu). .... SETI uses a large number of Internet-connected computers —most of them desktop PCs—to download and analyze radio telescope data, and to upload the results during idle times." [20]
In a telemedicine environment, it is easy to see where this tremendous distributed super computing capable operation could benefit all. One concept of a grid computing applications that could work assuming security and privacy can be maintained involves the establishment of a national Healthcare grid. Sharing of one major national grid to support hospitals, small rural and urban medical facilities and offices that need to be in the loop, emergency healthcare facilities, and more would lead to a cost sharing that potentially could be the basis for a survivable system for the nation that supports not only healthcare but concerns for our nations vulnerability from terrorist

There currently exist a "Smallpox Research Grid" uses a SETI-like model to analyze interactions between virus protein targets and a catalog of tens of millions of drug molecules (http://www.grid.org/projects/ smallpox).[20]

In addition to the SETI-like grids, a host of grids are currently interconnected through a maze of Universities across the nation being designed to study, along with literally thousands of smaller problems, the phenomena of gravitational waves. This Grid structure called "Medusa" is a large Beowulf-class parallel computer configuration, built and operated by the LIGO Scientific Collaboration (LSC) group at the University of Wisconsin - Milwaukee (UWM). The system went operational in August 2001 and is being used to develop and prototype data analysis for the Laser Interferometer Gravitational-wave Observatory (LIGO). Medusa can be used to support other problems now that it is operational but in the beginning it was and tends to be exclusive to Gravitational wave problems.[21] There are over 300 nodes in Medusa and it offers over 23 Terabytes of disk storage for its users.

A wide variety of commercially available grids are available if processing is required to support our telemedicine needs. Among these are the following (not all inclusive):

• Access Grid (http://www.accessgrid.org) An ensemble of resources to support group-to-group interactions across a grid.

• Computer Associates (http://www3.ca.com/Solutions/Collateral.asp?CID=52328&ID=2835): Managing on-demand computing.
• Fujitsu (http://pr.fujitsu.com/en/news/2002/04/22.html): Grid solutions for the sciences.
• HP (http://h71028.www7.hp.com/enterprise/cache/6842-0-0-0-121.aspx):Adaptive enterprise.
• IBM (http://www.ibm.com/grid): Grid applications for vertical markets, including banking and financial, life sciences, automotive, aerospace, chemical, electronics, petroleum, education, and government. (potentially acceptable for medical data analyses)
• Sun (http://www.sun.com/solutions/infrastructure/grid): Sun infrastructure solution for grid computing.

Grid Security

Finally, one of the larger issues with grid computing as with other aspects of the telemedicine activity is the consideration of grid computing security features. Ramakrishnan addressed this aspect of security for the next generation of grids in a recent article.[22] In his article Rama noted that in early generation grid structures security was not as serious a problem because the systems generally were taken across mutually trusted systems usually owned within the same enterprise or company. However, with the linkages now operating through a maze of communications links some of which can be Internet service arrangements, the management of security can be a nightmare of anguish for the manager. To address this the Grid Forum (a committee of Grid users and providers that was formed to help solve some of the problems and to address some of the issues associated with various grid designs) has formulated a set of fundamental security building blocks that are designed to offer a secure capability when interacting within a specified grid structure or set of designed structures. The most popular of the set of blocks identified is referred to as the Grid Security Infrastructure (GSI)[23]. This infrastructure, provided as part of a toolkit package (http://globus.org/toolkit) provides secure authentication and certification using the standard Public Key Infrastructure (PKI) process that most major telecommunications systems now are employing. There are at a number of Web Service security specifications under development (referred to as WS- Security Specs – http://www-106.ibm.com/developerworks/webservices/library/ws-secure) with participation from IBM, Microsoft, and Verisign among others. These specifications reference specific programming language sets to be used, protocols, and applicable means for monitoring and evaluating security in a grid environment. Rama noted in his concluding remarks that given the trend to work towards international grid designs, that countries will be required to resolve cross-nation issues such as cost sharing, legal data sharing issues, and more to support next-generation opportunities and laws will have to be addressed to help resolve disputes both relative to security as well as performance. For the telemedicine environment, which is international in scope, this may mean an even more rigid structure of protection mechanisms will be needed. Today's operations are not sufficient as yet but it the issues are being addressed.

Knowbots

Knowbots (or Knowledge Robots) are intelligent software agents that automate the repetitive tasks of humans. Knowbots have been found to be positively associated with higher learner completion rates in the workshops. In addition, knowbots learning-support tools aimed at reminding cognitively impaired patients of functions and activities that required attention were found to be as effective in motivating a patient to perform the function as human intervention.

Research performed over the past decade using Knowbots has suggested that the application of agent intelligent technology to online learning holds promise for improving completion rates, learner satisfaction, and motivation.[24]

There are many challenges facing the application of knowbots comprising Asynchronous Learning Networks (ALN) components. From our referenced research on the subject of ALN a number of interesting findings have been uncovered. [24]

In classroom-controlled environments there is as preponderance of data suggesting that the potential for losing cohesiveness and spontaneity of a classroom experience occurs when an ALN element is introduced as the learning mechanism. It was found that the instant availability of a human tutor online would be ideal. However, the cost and availability were shown to be limiting factors in providing this capability is therefore was no more realistic than continuously providing a human tutor for the traditional classroom. Often students simply wanted questions answered - and, it was indicated that they would be happy with any type of effective immediate feedback – human or machine.

> "An intelligent agent is any program that can be considered by the user to be acting as an assistant or helper, rather than as a tool in the manner of a conventional direct manipulation interface. An agent should as well display some, but perhaps not all, of the characteristics that are associated with human intelligence: **learning, inference, adaptability, independence, creativity, etc."** [24]

When initiating research using intelligent agents in a medical environment (e.g., cognitive studies of patients with some levels of Dementia or Alzheimers) a number of characteristics or qualities of the robotic instrument being introduced should be considered. These include: [25]

- Autonomy: An Intelligent Agent initiates and exercises control over its own actions in a Goal-oriented manner - - the agent must accept high-level commands indicating what a human wants and must decide how and when to satisfy the requests.
- Collaborative: An Intelligent Agent must not blindly obey commands but can offer alternatives and request clarification, or even refuse to perform a function in some instances
- Flexible: An Intelligent Agent's actions should not be scripted but rather the agent should exhibit a capability to dynamically choose actions to invoke and their sequence based on response to the state of its external environment.
- Self-starting: unlike standard computer programs directly invoked by the user, an Intelligent Agent must be capable of sensing changes in its environment and decide when to act.
- Temporal continuity: An Intelligent Agent is a continuously running process, not a one-shot computation that maps a single input to a single output and then terminates.
- Personality: An Intelligent Agent must exhibit a well-defined believable personality that facilitates interaction with human users.
- Communication ability: An Intelligent Agent should be able to engage in complex communication with other Intelligent Agents including people, to obtain information or enlist help to accomplish its goals.

- Adaptability: An Intelligent Agent must be capable of automatically customizing itself to the preferences of its user on the basis of previous experience. It also must automatically adapt to changes in its environment.
- Mobility: An Intelligent Agent must have a capability to transport itself from one machine to another and across different system architectures and platforms.

It is a given that no single Intelligent Agent (referred to also as Knowbots) would necessarily have all these characteristics. However, there are a number of robotic systems developed in current research programs that approach these characteristics and thus show promise for the future that systems such as these will exists and will be a truly representative robotic tool supporting future telemedicine applications as we have envisioned. Etzione and Weld both agree "...There is little agreement about the relative importance of different properties, but most researchers agree that these are the characteristics that differentiate agents from single programs." [26]

Finally, I would like to offer some examples of work currently in progress that utilize the aspects of wireless, nanoscience and technology, and knowbots Intelligent Agent technology.
Table 3 synopsizes a set of current robotic-like systems that are being applied in experimental environments to assist the aging population in developing their motor skills and in some instances serving to support the cognitive capabilities of an elder individual.

Table 3. Summary of Current Intelligent Agent Research Products.

ITEM	DESCRIPTION	FUNCTION	REFERENCE
1	Robotic Wheelchair	Outdoor Navigation	Yanco[27]
2	NAVCHAIR	Navigation System	Levine[28]
3	Gesture Pendant	Computer Vision System for Home Automation Control and Medical Monitoring	Starner[29]
4	Digital Family Portraits	Patient Activity Web Available Monitoring System	Mynatt[30]
5	Cueing Device	Assist Dementia patients	Mihailidis[31]
6	NeuroPage©	Memory Aid	Wilson[32]
7	Autominder	Intelligent Cognitive Orthotics system	Pollack[33]
8	Robotic Assistant	Robotic devices for Nursing Homes	Pineau[34]
9	Robotic Walker	Walker Guidance System	Morris[35]
10	Activity Compass	Intelligent Direction Support	Patterson[36]

Summary

Andy Grove, CEO of Intel said it best:

> *"[Healthcare] is the largest segment of the economy in the U.S. and...it is becoming too expensive to deliver. We're still living in the "mainframe" era of healthcare...[We] can't, as a society, afford to devote any more of our economy to it...[What]we need is ...the healthcare equivalent of the low cost PC."*[37]

The preceding discussion of wireless as it exists today, its security issues and solutions, its applicability to healthcare and support for lowering cost of healthcare through telemedicine, together with the strides being taken to introduce nanotechnology, grid computing, and Knowbotics all represent strides towards that proverbial low cost PC. With thought and attention to the future offerings of technology, telemedicine will succeed in being a major part of the solution that we in this nation have been looking for. In the past, the price to implement a telemedicine operation has proven too costly to go forward, but with today's lower cost wireless and improved data transfer rates, shared grid service arrangements, and introduction of intelligent agents it is no longer a losing proposition and deserves our focus and drive to make it work.

References

1. Ackerman, Michael, R. Craft, F. Ferrante, M. Kratz, S. Mandil, H. Sapci, "Telemedicine Technology," National Library of Medicine, Telemedicine Journal and e–Health, Chapter 6, Volume 8, Number 1, 2002.

2. *Conference Goal*, 2nd Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine and Biology, 2002-05-02 - 2002-05-04, Madison, Wisconsin; http://www.memsnet.org/events/979016400-9/

3. Fowler, Thomas B., "Cheap Internet Access: The Emergence of Bandwidth on Demand and the Changes it Will Bring tot eh Telecommunications Infrastructure," *The Telecommunications Review,* Mitretek Systems, Inc., MP96W0000139, Vol. 7. 1996.

4. IEEE Std 802.11™-1997, *Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, 1997.*

5. IEEE Std 802.11g™-2003, *IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band,* [Amendment to IEEE Std 802.11™, 1999 Edition (Reaff 2003), as amended by IEEE Stds 802.11a™-1999, 802.11b™-1999, 802.11b™-1999/Cor 1-2001, and 802.11d™-2001]

6. HIPAA, Department of Health and Human Services, Office of Secretary, Health Insurance Reform: Security Standards, Centers for Medicare & Medicaid Services (CMS), 45 CFR Parts 160, 162, and 164, [CMS-0049-F], RIN 0938-AI57, 2003.

7. Davie, Bruce, et al, "Bringing Health-Care Applications to the Internet," *IEEE INTERNET Computing,* May-June 2001; http://computer.org/internet/ 1089-7801/01/

8. Tanenbaum, Andrew S., Computer Networks, Prentice Hall PTR, Fourth Edition, 2003, pp.781-783.

9. Fluhrer, S., et al, "A Weakness in the Key Scheduling Algorithm of RC4," *Proc. of the Eighth Annual Workshop on Selected Areas in Cryptography*, 2001.

10. Fout, Tom, *Making IEEE 802.11 Networks Enterprise-Ready,* **http://www.microsoft.com/technet/treeview/default.asp?url=/TechNet/itsolutions/network/deploy/depovg/ieee802.asp** , Microsoft Corporation, April 2001.

11. Emerich, Dwaine F. and Christopher G. Thanos, "Nanotechnology in Medicine," *Expert Opinion Biol. Ther.*, Ashley Publications, 3(4), 2003.

12. Pentland, Alex (Sandy), "Healthwear: Medical Technology Becomes Wearable," *Computer Magazine*, Vol. 37, No. 5, May 2004, pp. 42-49.

13. Sairam, K.V.S.S.S.S, et al, "Bluetooth in Wireless Communications," *IEEE Communications Magazine*, June 2002.

14. Bluetooth Specification, "Specification of the Bluetooth System – Wireless Connections Made Easy," version 1.1, Feb 22, 2001. (also available at http://www.bluetooth.com)

15. Feynmar, R., "There's Plenty of Room at the Bottom." Presented at the *American Physical Society* Annual Meeting, Engineering and Science, Caltech (1959).

16. Freitas, Robert A., Jr., "The Future of Nanofabrication and Molecular Scale Devices in Nanomedicine," *Future of Health Technology*, Edited by R.G. Bushko, IOS Press, The Netherlands, 2002, pp. 45-59.

17. Dishman, Eric, Quoted from Verbal presentation : "Testimony to U.S. Senate Special Committee on the Aging," Director and Senior Research Scientist, Intel, April 27, 2004, Dirksen Senate Office Building, Washington, D.C

18. Dishman, Eric, "Inventing Wellness Systems for Aging in Place," *Computer Magazine*, IEEE Computer Society, May 2004, pp. 34-41.

19. Pollack, Martha E., "Active Technology for Aging Population," Presented at the Special Committee Hearing on Aging for the U.S. Senate, April 27, 2004.

20. "Grid Computing 101: What's All the Fuss About?" Compiled by: IT Professional's editorial board, *IT Professional Magazine*, IEEE Computer Society, March-April, 2004, pp 25-33.

21. Welcome to Medusa Website:
http://www.lsc-group.phys.uwm.edu/beowulf/medusa/index.html

22. Ramakrishnan, Lavanya, " Securing Next-Generation Grids," *IT Professional Magazine,* IEEE Computer Society, Vol. 6, No. 2, March –April 2004, *pp. 34-39.*

23. Foster, I, et al, " A Security Architecture for Computational Grids," *Proc. of Fifth ACM Computer and Commun. Security Conf.,* ACM Press, 1998, pp. 83-93.

24. Thaiupathump, Choonhapong, et al, "Intelligent Agents for Online Learning," *Journal of the Asynchronous Learning Networks*, Volume 3, Issue 2, November 1999.

25. Selker, T., "Coach: A Teaching Agent that Learns," *Communications of ACM*, Vol. 37. No. 7.

26. Etzioni, O., and Weld D. S., "Intelligent Agents on the Internet: Fact, Fiction, and Forecast," *IEEE Expert*, pp. 44-49, August 1995.

27. Yanco, H.A., "Development ad Testing of a Robotic Wheelchair System for Outdoor Navigation," *Proc. of the 2001 Conf. On Rehabilitation Engineering and Assistive Technology*, 2001.

28. Levine, S. P., et al, "The NavChair Assisted Wheelchair Navigation System," *IEEE Transactions of Rehabilitation Engineering*, 7(4), 1999, pp. 443-451.

29. Starner, T. et al., "The Gesture Pendant: A Self-Illuminating, Wearable, Infrared, Computer Vision System for Home Automation Control and Medical Monitoring," *Proc. of IEEE Intntl. Symp. On Wearable Computing*, 2000, pp. 87-94.

30. Mynatt, E.D., et al., "Digital Family Portraits: Providing Peace of Mind For Extended Family Members," *ACM Conference on Human Factors in Computing Systems*, 2001, pp. 333-340.

31. Mihailidis, A., et al, "The Development of a Computerized Cueing Device to help People with Dementia be More Independent," *Technology And Disability*, 13(1), 2000, pp. 23-40.

32. Wilson, B.A., et al, "Evaluation of NeuroPage©:A New Memory Aid," *Journal of Neurology, Neurosurgery, and Psychiatry*, Vol 63, 1997, pp. 113-115.

33. Pollack, M.E., et al, "Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment," *Robotics and Autonomous Systems*, 44(3-4), 2003, pp. 273-282.

34. Pineau, J., et al "Towards Robotic Assistants in Nursing Homes: Challenges and Results," *Robotics and Autonomous Systems*, 42(3-4), 2003, pp. 271-281.

35. Morris, A., et al, "A Robotic Walker that Provides Guidance," *Proceedings of the IEEE International Conference on Robotics and Automation*, 2003.

36. Patterson, D., et al, " The Activity Compass," *Proc. of the 1st International Workshop on Ubiquitous Computing for Cognitive Aids*, 2002.

37. Schlender, B., "Intel's Andy Grove: The Next battles in Tech," *Fortune*, May 12, 2003, pp. 80-81.

Bibliography

[BARA98] Bisain, Abhijeet S., *Intelligent Network Based Wireless Services for Telemedicine*, Te#B9127University of Maryland, MS Thesis, 1998: *[http://www.imse.med.tu-muenchen.de/mi/papers/ieee98/ieee98.html]*

[CHIM97] Chimiak, William J., et al, "An Architecture for Naval Telemedicine," *IEEE Transactions on Information Technology in Biomedicine,* VOL. 1, NO. 1, March 1997.

[ENGE98] Engelmann, Uwe, et al, "A Three-Generation Model for Teleradiology," *IEEE*

Transactions on Information Technology in Biomedicine, VOL. 2, NO. 1, March 1998.

[GLOM99} Glombitza, Gerald, et al, "Virtual Surgery in a (Tele-) Radiology Framework," IEEE *Transactions on Information Technology in Biomedicine* VOL. 3, NO. 3, September 1999.
[HERN01] Hernández, Alfredo I., et al, "Real-Time ECG Transmission Via Internet for Non-clinical Applications," *IEEE Transactions on Information Technology in Biomedicine,* VOL. 5, NO. 3, September 2001.

[HOLL99] Holle, Rolf, Gudrun Zahlmann, "Evaluation of Telemedical Services," *IEEE Transactions on Information Technology in Biomedicine,* VOL. 3, NO. 2, June 1999.

[ISTE99] Istepanian, Robert S. H., "Telemedicine in the United Kingdom:
Current Status and Future Prospects," IEEE *Transactions on Information Technology in Biomedicine,* VOL. 3, NO. 2, June 1999.

[ISTE00] Istepanian, Robert S. H., "Guest Editorial: Special Issue on Mobile Telemedicine and Telehealth Systems," *IEEE Transactions on Information Technology in Biomedicine,* VOL. 4, NO. 3, September 2000.

[LAMM01] Lamminen, Heikki, et al, "A Three-Year Follow-Up of Finnish Telemedicine Programs," *IEEE Transactions on Information Technology in Biomedicine,* VOL. 5, NO. 2, June 2001.

[LAUC02] Lau, Christopher, R. et al, "Asynchronous Web-Based Patient-Centered Home Telemedicine System," *IEEE Transactions on Information Technology in Biomedicine*, VOL. 49, NO. 12, December 2002.

[LUND04] Lundberg, Lydia, *Assisted Technology for Aging Population*, Testimony to US Senate Special Committee on Aging, April 27, 2004.

[MURA94] Murakami, Hajime, et al, "Telemedicine using Mobile Satellite Communications," *IEEE Transactions on Information Technology in Biomedicine*, VOL. 41, NO. 5, May 1994.

[REIN00] Reina-Tosina, Javier, Laura Roa, and Manuel Rovayo, "NEWBET: Telemedicine Platform for Burn Patients," *IEEE Transactions on Information Technology in Biomedicine, VOL. 4, NO. 2, June 2000.*

[SHEY02] Sheybani, Ehsan O. and Ravi Sankar, "ATMTN: A Telemammography Network Architecture,"" *IEEE Transactions on Information Technology in Biomedicine*, VOL. 49, NO. 12, December 2002.

[SIEM02] TD-SCDMA@vantage, "Going 3G and beyond," Mobile Business, A50001-N3-P82-1-7600.

[TAKI 01] Takizawa, Masaomi, et al, "Telemedicine System Using Computed Tomography Van of High-Speed Telecommunication Vehicle," *IEEE Transactions on Information Technology in*

Biomedicine, VOL. 5, NO. 1, March 2001

[VARA02[Várady, Péter, et al, "An Open Architecture Patient Monitoring System Using Standard Technologies," *IEEE Transactions on Information Technology in Biomedicine*, VOL. 6, NO. 1, March 2002.

[WOOD01] Woodward, B., R. S. H. Istepanian, and C. I. Richards, "Design of a Telemedicine System Using a Mobile Telephone,"" *IEEE Transactions on Information Technology in Biomedicine,* VOL. 5, NO. 1, March 2001.

[ZHAN00] Zhang, Jianguo, et al, "Real-Time Tele-consultation with High-Resolution and Large-Volume Medical Images for Collaborative Healthcare," *IEEE Transactions on Information Technology in Biomedicine,* VOL. 4, NO. 2, June 2000.

Towards Technical Interoperability in Telemedicine

Richard L. Craft
Sandia National Labs
rlcraft@sandia.gov
http://telemedicine.sandia.gov

Abstract

In order for telemedicine to realize the vision of anywhere, anytime access to care, the question of how to create a fully interoperable technical infrastructure must be addressed. After briefly discussing how "technical interoperability" compares with other types of interoperability being addressed in the telemedicine community today, this paper describes reasons for pursuing technical interoperability, presents the "Telemedicine System Interoperability Architecture" (a proposed framework for realizing technical interoperability), identifies key issues that will need to be addressed if technical interoperability is to be achieved, and suggests a course of action that the telemedicine community might follow.

Keywords: Telemedicine, Interoperability, Architecture, Design, Protocol, Technology, Device

Introduction

During the last decade, as practitioners in telemedicine were focusing on clinical efficacy, patient satisfaction, and increased access to care, economists were exploring the business case for over-the-wire approaches to care delivery, and lawyers and policy makers were developing a financial and regulatory environment in which telemedicine practice could thrive, engineers were functioning largely as system integrators – cobbling together off-the-shelf components to create one-off infrastructure designs that served as vehicles for the activities of the clinical practitioners. While this was necessary given the nascent nature of these other activities, there is more that must be accomplished from an engineering perspective if telemedicine is to mature as an industry and if telemedicine-based practice is ever to reach its full potential. One of the most important of these things is the development of interoperability approaches for telemedicine.

The purpose of this paper is to address one aspect of interoperability: how elements in a telemedicine infrastructure interact with one another. The centerpiece of this paper is a discussion of the "Telemedicine System Interoperability Architecture" (TSIA), a proposed approach for achieving interoperability in telemedicine devices. This architecture is designed to support telemedicine-based operations as they are practiced today but is also intended to make it possible to implement a range of other care delivery structures.

This paper is organized into four sections. Within the telemedicine community, the term "interoperability" is used to mean different things. The first section of this paper briefly discusses different aspects of interoperability in telemedicine and identifies which kind of interoperability is being addressed in this paper and which kinds are not. The second section addresses the factors that influenced the design of the TSIA. The third section presents the Telemedicine System Interoperability Architecture, describes what this design allows that current approaches cannot, and discusses how this design addresses each of the influences described in the second section. The final section suggests specific objectives that the telemedicine community pursue in realizing technical interoperability and proposes a course of action for accomplishing this.

Aspects of Interoperability

The term "interoperability" is not used uniformly within the telemedicine community. There are at least three variations in usage worth considering. The first of these is "clinical interoperability". Different healthcare systems have different ways of doing things. The existence of this "culture" in any such system is important inasmuch as an understanding of what is expected in any given situation can help that culture operate more efficiently than would be possible if every intention and every decision had to be communicated in the moment. Telemedicine-based care delivery systems can transcend these organizational boundaries such that individuals belonging to different healthcare communities interact to accomplish some joint care delivery objective. In these settings, establishing "clinical interoperability standards" that clearly articulate expectations regarding what tasks will be accomplished by whom in a given kind of situation and what

will be communicated and when helps the distributed team operate as whole without a lot of unnecessary delays or interactions.

A second variation, "information interoperability", is the focus of much work underway today in the medical standards communities (e.g., HL7 [1] or SNOMED CT® [2]). The issue here is ensuring that interacting parties are capable of sharing information accurately and efficiently. These efforts typically emphasize the development of a common concept base, a common vocabulary for representing that space of concepts, and a common syntax and semantics for communicating these concepts back and forth.

The third variation is "technical interoperability". This can be divided into two parts – communications interoperability and device interoperability. In communications interoperability, the chief is concern is on how to move information from one place to another. Problems such as mismatched communications equipment [3] and network quality of service (e.g., assuring that error rates do not exceed a given level or that data packets take no more than a certain amount of time to move between stations) are included in this part. In device interoperability, the emphasis is on how the various components that make up a station interact and on the mechanisms by which communicating stations interact. This paper is concerned with this latter portion of technical interoperability.

What Motivates This Effort

Three sets of influences played a dominant role in shaping the Telemedicine System Interoperability Architecture. These were:

- needs in current telemedicine,
- trends in technology, and
- changing employment concepts.

Needs In Current Telemedicine

As an industry, telemedicine is young. Vendors who have survived to date have done so either by building custom systems for specific users operating in specific settings, by creating systems that are designed to support a given clinical need (e.g., management of congestive heart failure), or some combination of the two. In either approach, the lion's share of their business is built around integrating disparate components into tightly coupled systems that deliver certain capabilities to clinical users. Customers who buy these systems typically have a specific clinical application in mind. They buy a system in order to employ a particular set of clinical capabilities and typically have a specific population of users (both clinicians and patients) in mind when they do so.

While this approach to life makes sense in light of telemedicine's recent history of government-funded studies meant to validate its worth, it introduces some side effects that are not so desirable if telemedicine is to move from research or employment on the fringes of healthcare to a central element of general care delivery. In particular, the current approach to telemedicine:

- creates systems that lack open connectivity,

- limits both consumer and vendor autonomy,
- reduces telemedicine's relevance to mainstream care delivery, and
- leads to higher than necessary infrastructure costs.

A Lack of Open Connectivity

Today's telemedicine systems are like intercoms, not telephone networks. They allow individuals within a closed network of users to interact with one another but not with individuals in other networks. In part, this is an artifact of a telemedicine community mindset that "this is how telemedicine is done" and, in part, it is owes to the fact that vendors, lacking time or funds to address interoperability, have yet to work out the details of how to make their stations communicate with one another. Irrespective, the result is that telemedicine systems are less useful than they might be if any station created by any vendor could communicate with any other station created by another other vendor.

Limited Consumer and Vendor Autonomy

Because telemedicine systems lack open connectivity consumers must either buy their systems from a single vendor or pay a developer to integrate different systems into an interoperable network. The first option presents the purchaser of these systems with at least three problems. First, the vendor may not be able to the deliver all of the capabilities that the purchaser needs over the life of the telemedicine systems operation. If additional capabilities are needed beyond what the vendor has already developed, there can be significant time and money involved in getting these features fielded. Second, fielding a telemedicine network can involve a substantial investment. The effect of this investment can be to leave the customer "captive" to the vendor. Even if a current vendor proves unsatisfactory (e.g., unable or unwilling to deliver a needed capability), a customer may be unable to pay what is required to switch to a new vendor. Third, this young industry has seen a number of companies – some large and well-known – enter the telemedicine marketplace only to later withdraw. Given this a customer considering investment in a telemedicine network will have to think long and hard about whether or not particular candidate vendors are likely to be around to service customer needs after the sale has been completed. Taken together, these factors have to cause potential users of telemedicine-systems to question whether or not the industry has advanced to a state where an investment in telemedicine technology is a sound use of resources.

None of this is to say that the vendors hold all of the cards in the telemedicine marketplace. Since telemedicine systems are typically sold as turnkey capabilities, being a player in the telemedicine marketplace typically means having to be a jack of all trades. While a vendor's strength may be in a particular technology domain (e.g., patient records or a given kind of medical instrument), the vendor still has to deliver all of the other capabilities that go into a telemedicine system (networking, application programming, user interface development, etc.). This fact can present a significant barrier to entry for those vendors considering building a business in the telemedicine arena. For those already in business, this way of operating can divert funds that might otherwise be invested in developing better capabilities in the vendor's area of expertise. Either way the net effect is a greatly reduced suite of choices than might otherwise be available in the telemedicine marketplace.

Reduced Relevance to Mainstream Care Delivery

In their book, Telemedicine and the Reinvention of Healthcare [4], Marc Ringel and Jeffrey Bauer state:

> Telemedicine, one of the major forces shaping the future of healthcare, is widely misunderstood. Its long-term impact on healthcare is obscured by excessive concerns with short-term policy problems, a misleading focus on narrow definitions, or utopian expectations of technology. People who overreact to telemedicine's early difficulties or underestimate its scope will be surprised by its real power. *Telemedicine will ultimately revolutionize healthcare – restructuring virtually every relationship and activity that define late twentieth century medicine.*" [emphasis added]

That this is so seems clear to the author. It is not a matter of if it will happen but when.

But that's the future. Today, in talking with doctors and nurses who do not practice using telemedicine-based techniques, it is not uncommon to hear them say that such approaches are an interesting curiosity but not of any real use in mainline care delivery settings. In part, this point of view reflects a short-sightedness on the part of these clinicians that is rooted in the belief that how medicine is done today in "standard" settings is how it *should* be done. Interestingly, this same belief seems to be an implicit part of the telemedicine community's own culture. It is almost impossible to go to a telemedicine conference or other meeting on telemedicine without hearing that telemedicine is what you do for "those folks out there who don't have access to 'regular' care delivery mechanisms." Using telemedicine-based approaches for "everyday care delivery" is not yet part of the telemedicine community's rhetoric. This is unfortunate given that one of the persistent problems in telemedicine has been the community's heavy dependence on government funding for survival. As will be discussed below, making telemedicine-based care delivery self-sustaining will require broadening the number of places in which it is used and the range of people who can use it. If it hopes to mature, the telemedicine community will eventually have to expand its charter to include servicing not just poor, distant communities but also affluent neighborhoods located just a few miles from where the caregivers operate.

In addition, telemedicine will need to broaden repertoire of capabilities. Today, the number of clinical processes that can be carried out over-the-wire is exceedingly small compared to the number of things that can be done in traditional face-to-face settings. Even if doctors and nurses operating in mainline settings were inclined to embrace computer-mediated care delivery techniques, it is very likely that most would not able to field a rich enough suite of capabilities to make their transition to telemedicine worthwhile.

One additional change that will be required to make telemedicine relevant to mainline care deliver is the development of telemedicine systems that are easily customized to meet site- and situation-specific operational requirements. Today's telemedicine systems are generally built for a specific setting or a specific disease state. One of the ways that this expresses itself is that telemedicine stations are frequently designed with specific connectors for specific devices in their equipment suite (i.e., "the scope connects here;

the pulse-oximeter goes here; ..."). Nothing else other than the intended devices can attach to those connectors. Even if the station designers have been clever enough to avoid the connector trap, the software that recognizes which devices are attached to the station may only know about the devices that the station was engineered to handle. In both cases, adding new capabilities to the station can require a significant investment of money and engineering time. If the telemedicine community wants to move telemedicine into the mainstream, telemedicine stations and the "peripherals" that they support will need to enable the kind of plug-and-play integration and operation that typifies today's audio-video and consumer computer markets; however, doing this will fundamentally change the telemedicine industry. While system integration groups will still exist, the shift toward end-user integration of stations will mean that the majority of companies in the industry will focus on development of new or more cost effective capabilities.

Higher Than Necessary Infrastructure Costs

A final side-effect of today's way of doing business in telemedicine is that telemedicine systems are more costly than they would be in a different kind of world. There are several factors that contribute to this. First, telemedicine has not yet transitioned to a commodity-style business. Today's practice of customizing telemedicine stations for specific operational needs translates into engineering time. As production runs typical of today's telemedicine marketplace are relatively small, there are fewer units over which to amortize engineering costs for systems are more expensive than the would be in a commodity-oriented environment.

Next, today's systems are meant to function as turnkey capabilities. They can be carried into a home, connected to the phone line, and be ready to operate. Because of this, telemedicine stations are self-contained units that include all of the parts that are needed to deliver needed functions. This can include displays, interface devices, processors, data storage devices, modems and/or networking interface cards, etc. While currently necessary, this approach is more costly than one in which required functions can be shared with other applications (e.g., in a home setting a television might be used as an interface device in lieu of a flat-panel display that would otherwise need to be integrated into the station).

In the end, equipment and operation-related costs (e.g., network connectivity) will factor into decisions on the part of clinicians regarding whether or not to embrace telemedicine-based approaches in lieu of traditional face-to-face approaches. In addition, if the issue of cost is addressed, a world in which patients and clinicians alike purchase their own stations for their own use (much as everyone today buys their own telephone) is easy to imagine.

Trends in Technology

The second set of considerations that shaped the Telemedicine System Interoperability Architecture was the direction in which computing and networking seemed to be heading. Dominant considerations here included:

- growth in networking,
- grid computing / web services,

- micro-/nano technology, and
- software agents and related technologies.

Growth in Networking

While not a new notion, the fact that the Internet continues to grow both in extent and in throughput is significant to telemedicine's future. During the last decade, telemedicine has focused heavily on servicing places where connectivity is sometimes in question and where bandwidth is often severely limited. This operational constraint has influenced how telemedicine systems are designed. For example, the emergence of "store-and-forward" techniques is an artifact of this world of limited communication resources. In contrast to this, some telemedicine studies have begun to consider what telemedicine-based operations might look like if bandwidth was essentially unlimited [5].

This kind of future wired world is not just theory. The number of end users with accessing to high bandwidth connections is growing. For example, the number of digital cable TV users in the United States has grown from 6 million in 2000 to over 21 million by the end of 2003 [6]. As of March, 2004, over 45% of U.S. homes possessed high-speed Internet access (cable or DSL) [7].

While some of the implications of more bandwidth (such as the ability to move more data – bigger images, etc.) have been discussed in the telemedicine community, others have not. One important example is the relationship between bandwidth and where system functions are performed within a system. As an example, consider the problem of designing an imaging capability into a telemedicine system. If bandwidth is low responsibility for producing a good image will very likely fall to the tech who runs the acquisition system. Images captured by the tech will be buffered on the acquisition station and one or a handful selected for transmission to the radiologist who will read the images. All processing of the image will probably be done on the reviewer's station where the radiologist is free to interact with the image. In contrast to this, high bandwidth communications offer the system designer more freedom. For instance, since images move "instantaneously" between acquisition system and reviewer station, the radiologist might actively engage in the image capture process. In addition, the processing algorithms used by the radiologist might be moved onto a mainframe located somewhere on the Internet. That these underlying architectural changes had occurred would be transparent to the users of the system but could significantly lower the costs of the associated stations.

Grid Computing / Web Services

As available bandwidth increases, it becomes possible to move computing capabilities away from a user's physical location and out onto the Net. Today this trend is taking shape in two domains: hardware and software. In the case of the first, the goal is to allow hardware assets spread across and network to be federated and made available on-demand for use by users on the network. When fully realized, this will mean that the assets on which a user draws need not be collocated with that user but spread out across the network that he accesses. In this way, relatively "weak" computers (like today's

PDAs) can invoke the capabilities of a virtual supercomputer to perform tasks for them [8].

On the software side, significant energy is going into the development of "web services". Unlike the traditional approach to software which delivers turnkey applications to the user's desktop, the main thrust in web services is to create software Legos™ that can describe their capabilities to other (often remotely located) Legos and can be "snapped together" on the fly to create larger software programs that deliver some desired service [9].

For telemedicine, these two trends point to a future in which telemedicine stations need not be rendered as monolithic pieces of equipment but can be assembled on the fly from a small suite of user-held hardware and software that dynamically makes use of "services" that exist "somewhere, out there" on the Internet.

Micro-/Nano Technology

One implication of the telemedicine community's goal of "anyone, anytime, anywhere" delivery of care is that it is possible to take the medical instrumentation associated with care to any place at anytime. Unfortunately, this is not yet possible. Engineers design equipment for use in certain settings. Much of the equipment that is integrated into telemedicine systems was never designed with telemedicine in mind. These devices are usually engineered for use in the relatively "benign" environments typical of clinical settings. They cannot handle temperature extremes, will break if dropped, etc. The size or weight of some of these devices can limit their mobility. Power requirements can chain them to the walls inside a facility.

Advances in micro- and nano-technologies suggest that there is the potential for recasting many of these clinical capabilities in forms that would allow them to be used in a much wider range of settings than is possible today [10]. Because these kinds of devices can be mass produced on the same scales as microelectronics, they also have the potential to significantly reduce the cost of the devices in which they are incorporated.

Software Agents and Related Technologies

The central idea behind this final technology trend is that some of the intelligent tasks currently performed today by humans will be done tomorrow by computers. The intention of these tools is to extend the reach of experts (e.g., by making it possible for these experts working in cooperation with these tools to process more things than could be done alone). The tools are also aimed at providing non-experts with the ability to effectively use knowledge is normally found only in experts. Larry Weed's Problem-Knowledge Couplers are a good example of both uses [11].

Changes in Employment Concept

The final influence that shaped the Telemedicine System Interoperability Architecture was a belief that changing demographics and similar factors would lead the telemedicine community into employment concepts other than those it has pursued to date. While this subject could be addressed at length, it seemed to the author that the aging Baby Boomers

and attendant rise in healthcare expenditures would drive significant restructuring in healthcare delivery processes. Simply denying benefits as a means of containing cost would not prove acceptable and means of establishing a more effective partnership between clinicians, patients, and their caregivers would be required.

Similarly, it seemed that lifestyle-based diseases would begin to cause problems for the healthcare community. Indeed, this has been borne out by the recent actions around the growing problem of obesity. Given this, it seemed that American healthcare would need to be reinvented to allow the clinical community to play a greater role in addressing the root causes underlying these problems.

As both kinds of pressures reside outside of the acute care world where "regular" medicine is practiced, the author assumed that telemedicine-based techniques could play a significant role in addressing these emerging needs. To do so, however, would require that telemedicine systems not be viewed as "durable medical equipment" prescribed for a period of time by as permanent fixtures in the lives of the patients with whom doctors and nurses are engaged. As such, telemedicine capabilities would need to fit more easily into every day life (e.g., not take up a big footprint on a table somewhere, not tether the patient to their home) and probably be treated as yet another appliance or suite of consumer electronics.

Finally, as noted above, the author believed that, if telemedicine is ever to become a robust industry, it would have to work to develop new, more profitable markets than those historically addressed. If a rich enough suite of capabilities could be developed and prices brought low enough, there was no reason that telemedicine-based methods could not be used to service patient populations in all economic strata of all locations – metropolitan and rural alike – in the country.

The Telemedicine System Interoperability Architecture

In light of these things, the Telemedicine System Interoperability Architecture was developed as a proposal for how the telemedicine community might achieve device-level interoperability. When the idea was birthed, some members of telemedicine vendor community found the notion of developing a detailed architecture that provided "the answer" threatening ("What if the standards that are produced make my current systems 'non-compliant'?" "What if compliance with the standards requires an investment that I can't afford to make?" "Will the standards threaten my ability to differentiate myself in the marketplace?" "Will they limit my ability to innovate or to adopt emerging technologies that customers are asking that I address?"). For this reason, TSIA was framed as an architectural *approach*. It was meant to be a starting point for debate within the community regarding what kinds of standards needed to be embraced. In addition to presenting this architectural framework, the full report [12] on this architecture identifies candidate technologies that might be selected by the community for use implementing various aspects of the architecture.

Why An Architecture And Not Just A List of Standards

One approach to specifying technical interoperability would be to simply enumerate a list of existing standards that stations must employ if they are to be considered interoperable. This approach works if the only consideration is station-to-station interoperability; however, more than this must be done if stations are to be readily customized by end users to meet a range of operational requirements.

The role of an architecture like TSIA is to define the logical blocks of functionality (record storage, information display, etc.) that make up a system, to specify how each unit of functionality can be physically embodied, and to articulate which logical and physical building blocks can be interconnected and how.

As an example, consider the personal computer world. A PC design typically includes a motherboard that serves to interconnect the rest of the computer's components and a suite of cards (e.g., networking or graphics) and other devices (hard drives, I/O devices, etc.) that deliver certain functions. Running on top of all of this is an operating system that mediates utilization of these resources by specific application software packages. In the PC world, end users are free to choose which mix of hardware and software they want to employ. This is all made possible by the existence of an overall architecture (both at the hardware and software levels) that allows different devices to interoperate even though created by different vendors and that allows different computers composed in this way to interoperate with one another. It should be noted that the existence of such architectures do not constrain vendors to fully implement every aspect of these architectures. For example, certain vendors exist today who make their living delivering turnkey, non-extensible systems that deliver a fixed suite of capabilities but are still able to interoperate in this world of computers. Likewise, while most of the computer vendors today sell these open architecture PCs, most of the units that they ship are factory-assembled systems built to customer specs and capable of being unboxed by their end users and run as-is without ever again being customized by their users.

The Architecture's Key Features

The Telemedicine System Interoperability Architecture embodies the following features:

- it is built around three sets of interfaces,
- it allows for plug-and-play operation of station peripherals,
- it is a component-based design capable of distributed, networked operation,
- it employs dynamic discovery and federation, and
- it is meant to integrate with other systems.

Three Sets of Interfaces

As a starting point, the TSIA assumes that there are three sets of interfaces within a telemedicine station that are important to consider (Figure 1). Station-to-station interfaces are those involved in allowing different telemedicine systems to interact with one another. Included in this set of interfaces are:

- videoconferencing,
- other forms of person-to-person communication,

- remote device monitoring and control,
- patient record exchange,
- session management (including network quality of service management)
- automatic station configuration,
- station location, and
- exploration and leasing of remote station services.

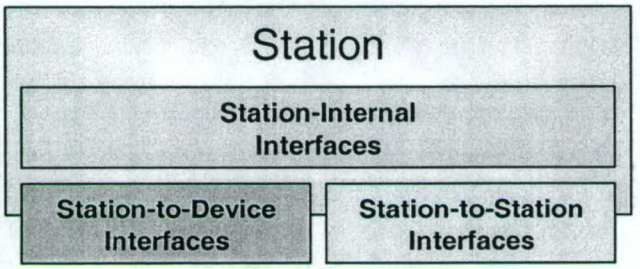


Figure 1. Three Sets of Interfaces

Station-to-device interfaces address interactions between a station and the various "peripherals" that can connect to it. Included in this list of peripherals would be things like medical instruments, data storage devices (e.g., patient record cards), and external user interface capabilities (such as might be supplied by a PDA or wireless tablet PC). Included in this set of interfaces are:

- general device plug-and-play operations,
- medical device monitoring and control,
- storage device interactions, and
- user interface device interactions.

Station-internal interfaces deal with how the various elements that make up a telemedicine station (e.g., displays, processors, record storage) interconnect and interoperate with one another. Included in this are:

- registration of components being added to a station,
- leasing of components and services for use in a given telemedicine session,
- dynamic interconnection of devices,
- on-the-fly customization of station configuration based on user preferences, and
- automated station self-configuration as required to support the introduction of new components.

These station-internal interfaces are included in the TSIA with a view to allowing telemedicine stations to be made of interacting components distributed across a network (Figure 2).

Plug-and-Play Operation of Station Peripherals

The proposed TSIA design allows for the attachment of three classes for station peripherals: storage devices, medical instruments, and user interface elements. Storage devices can be used to hold both medical records and "contexts" (discussed below) for

patients and/or clinicians. Medical instruments can include a range of devices capable of producing single-valued outputs (e.g., thermometers), continuous waveforms (e.g., ECG

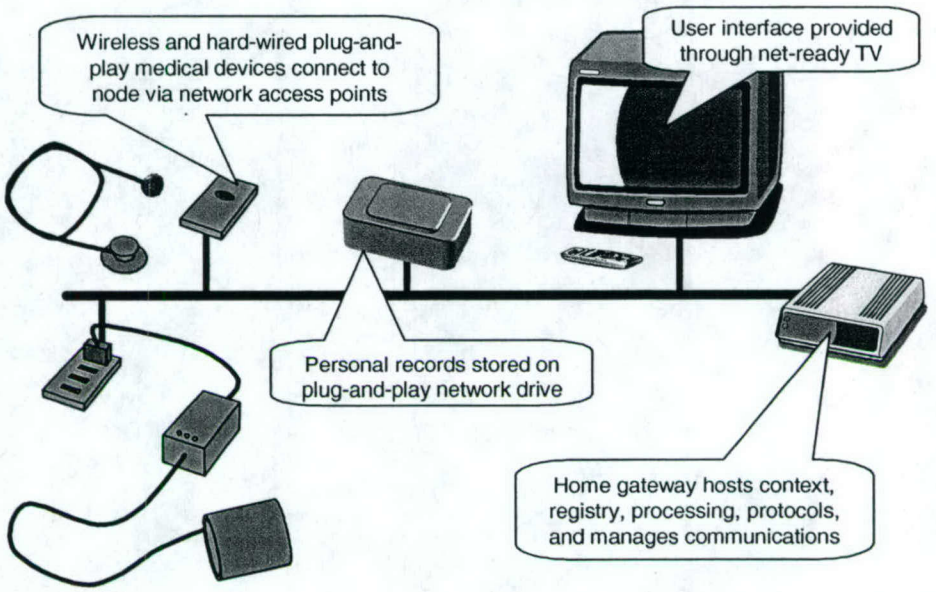


Figure 2. A Distributed Station Design

or stethoscope), and images (e.g., cameras or scopes). User interface elements can include simple text-based interfaces, graphical interfaces, voice/audio devices, etc.

For each such device, the architecture allows devices to be added and removed by end users at the time of use. If the station has never hosted a particular type of device (as determine by manufacturer ID and model number), the station has the ability to self-configure on the spot in order to automatically integrate the new device into the station. In like fashion, remote stations connecting to a station with capabilities that the remote stations have never before handled have the ability to configure themselves on the fly to make use of these capabilities.

Networked, Component-based Design

The Telemedicine System Interoperability Architecture allows for stations to be composed of distributed components. These components can all be located on a common network (such as the local area network that might exist in a home or in a doctor's office), as shown in Figure 2. The components that make up a station can also be remotely located. A telemedicine station's storage of patient records on remote record server, as shown Figure 3, is an example of this.

Dynamic Discovery and Federation

In TSIA, not only can station components be distributed across both local and wide area networks, but the exact suite of components that make up a "station" can change from session to session. Consider, the "telemedicine station" shown in Figure 4. In this example, the station is centered around a bed in hospital and contains several user

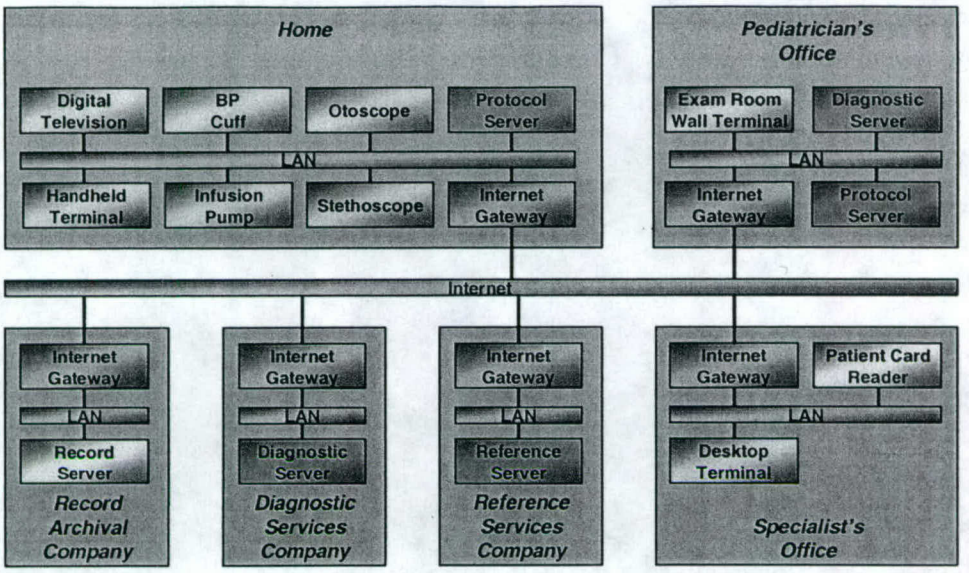


Figure 3. A Distributed Telemedicine System

interface devices (display, speaker, microphone) and an "access point" that serves as the interface for any new devices added to this "station". Depending on the needs of the patient occupying this bed, different kinds of devices, such as infusion pumps, imaging devices, and vital signs monitors can be dynamically added to and removed from the stations. Similarly, both clinicians and patients can identify services (e.g., patient record storage) that need to be attached to whatever station they are associated with at any given moment. The mechanism for doing this in TSIA is called a "context". This station component operates by specifying the location of service-oriented "components" that can be automatically integrated into a station on demand.

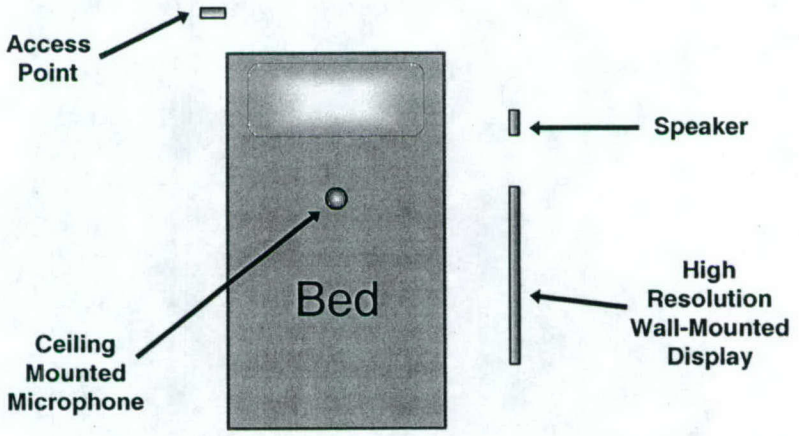


Figure 4. Hospital Bed as Telemedicine Station

Also, in a world of any station to any station operation, it cannot be assumed that two stations that are going to interact have a priori knowledge of each other's composition or capabilities. Given this, a central feature of the TSIA is the ability for remote stations to explore each other's capabilities and to "lease" these capabilities from each other on the

fly, even if the ability to use these capabilities has never before been present of the station making use of these remote features.

Integration With Other Systems

The Telemedicine System Interoperability Architecture is meant to allow telemedicine systems to integrate with other systems outside of the world of telemedicine. This expresses itself in at least two ways. The first is the ability of telemedicine stations to share the resources they employ with other kinds of systems. For example, in the distributed station design shown in Figure 2, the telemedicine station makes use of a television for its user interface component, of a home network as the "internal communications bus" that connected all of the station components to one another, and of a network-based hard drive as a storage location for all medical records generated by the telemedicine station. The inclusion of this kind of capability in the TSIA is central to being able to drive down the cost of telemedicine systems.

The second way that this integration expresses itself is in a station's ability to interoperate with other clinical systems. Today, this would mean primarily the ability to seamlessly move data between various kinds of storage systems (laboratory information systems, hospital information systems, etc.) and the ability to host certain capabilities (such as point-of-care lab instruments). As medical informatics matures, this will also mean things like the ability to dynamically engage "intelligent systems" (e.g., diagnostic software or workflow management systems).

The Station-Level Architecture

The Telemedicine System Interoperability Architecture is divided into two primary parts: those that deal with station-internal operations and those that deal with interactions between a station and external entities. Figure 5 depicts the station-level architecture.

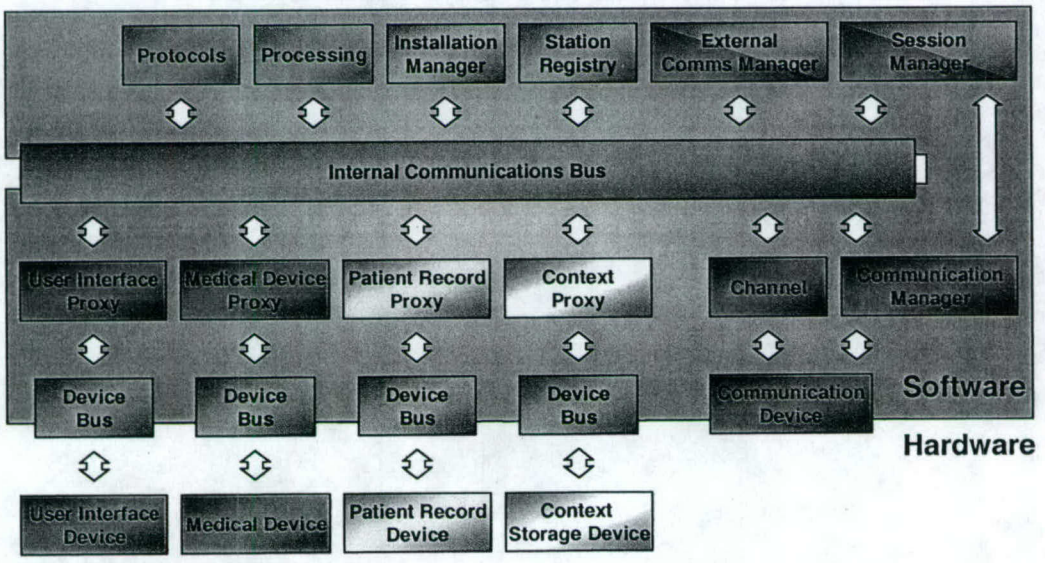


Figure 5. Logical Station Architecture

The components at the bottom left of the diagram represent *physical* resources that may be added to or removed from a station. These include ***user interface devices***, ***medical devices***, ***patient record storage devices***, and ***context storage devices***. Each of these devices connects to the rest of the station by means of a ***device bus***. This part of the station exists to support "hot plugging" of the devices into the station. The ***proxies*** that are shown between the internal communications bus and the device buses exist to encapsulate vendor-specific device communication behind standard ***interfaces*** that are understood by other system resources.

Communication devices can include things like the modem or Ethernet card in a PC that the station uses or a gateway device on a network that serves as the station's internal bus. In order to shield station components from the details of what kinds of devices exist to support external communications and how these devices are controlled, a ***communications manager*** is provided. Upon request, this software component establishes ***channels*** through which a station's components can connect to external resources. An ***external communications manager*** provides a single point of contact for mediating whether sessions will be established with remote stations. ***Session managers*** provide a means for external systems to "lease" a station's resources and, working with the communications manager, also monitor the status of external communication networks to ensure that quality of service requirements associated with interacting with remote resources are being met.

Within a station, communications between components are mediated by the ***internal communications bus***. A key feature of this bus is its ability to support quality of service requirements of the rest of the components that make up the station (e.g., it can deliver data from a medical device to a given user interface component in a specified amount of time).

The ***processing*** block in the diagram represents the data transformation components and other specialized building blocks (e.g., statistical analysis tools, "intelligent agents", person-to-person communication clients) that reside on the station.

In order to allow for automated downloading and installation of components onto a station, the ***installation manager*** is included in the station architecture. This component, working in conjunction with the device bus managers and the registry (described below), determines when a device that has just been added to station lacks the necessary infrastructure (i.e., proxy, processing, protocol, patient record, and user interface software) to function on that station and, when permitted by the station's operator, downloads from the web and installs the software that provides this infrastructure for the device.

Actual station configuration is controlled by three components that form the heart of the station. The ***station registry*** is used to catalog the list of components that the station can offer to support various clinical functions. As physical components are added to or removed from the station or as software components are added by the installation

manager, these changes are recorded by the station registry. ***Protocols*** are components that are responsible for intelligently acquiring resources needed to implement a clinical function, instructing these resources on how they are to interoperate, and then monitoring the operation of these resources for key events (e.g., a "leased" device being removed from the station prematurely or the operation terminating under normal conditions). As described above, ***contexts*** store information regarding patient and caregiver preferences and are responsible for configuring a station (or stations) in accordance with these preferences whenever a user logs onto a station. Every station will have a standard configuration that it uses in the absence of other configuration information. This is referred to as the ***default context***. ***Caregiver contexts*** and ***patient contexts*** can be introduced to the station via some form of ***context storage device***.

What's Different About This

So how does this differ from what is typically found today? First, everything in a TSIA station is a treated as a component. Nothing is connected to anything else until time of use. When the telemedicine station powers up, it takes stock of what components are available for use and then, based on information stored in the station's context, links together some or all of the available components to create the system that will be presented to the station's user. This is in contrast to today's designs in which all components are selected and interconnected at the time of design.

Second, every component in a TSIA station has a standardized logical interface that allows it to work with all other components with which it might every have a relationship. For example, the way that a blood pressure monitor operates (i.e., what commands it expects, what information it returns, and the way that this information is encoded) is specified in some sort of standard. All other elements of the station that would interact with the blood pressure monitor (e.g., user interface components used to monitor and control the device or elements of the patient record used to store readings generated by the device) would understand and employ this standard. Where physical components have an interface that departs from this, a proxy is used to handle translation between the two interfaces. In the case of a blood pressure monitor, a proxy might be used to convert a vendor-specific protocol for controlling and monitoring the device into the standardized protocol expected by other station components.

Third, communications mechanisms within the station are also standardized. The internal communications bus consists of both the media over which bits are moved and of a suite of services that allow components in the station to attach to "channels" as publishers or subscribers of various sorts of messages. As with other station components, these internal communication bus services present a standardized interface to the components that employ their services and all using components implement these same interfaces.

Fourth, device buses are used to standardize the way in which the station's peripherals attach to the station. Unlike current systems in which each peripheral may connect via a dedicated connector that is different than other device connectors on the station, the device bus allows any device to attach to any connector on the station and, depending on the device bus technology used, means that a station can host an arbitrary number of

peripherals. Example technologies that might work in this role are the USB standard that is now common in PCs, IEEE 1394 (commonly used in video equipment), or BlueTooth (the wireless peripheral interconnect standard).

Next, standardizing on a device bus approach means that medical instruments and other peripherals will need to include the appropriate connectors and medium for the bus and that these devices possess some degree of intelligence (Figure 6). At a minimum, peripherals in TSIA must be able to uniquely identify themselves (e.g., by manufacture ID, model number, and serial number). In the figure, such a device would be called a "self-aware device". Today, most station peripherals are "dumb devices". They can only respond to commands sent them and cannot identify themselves in a standard way to a station. For completeness, the figure shows two other possible configurations for devices that can attach to a TSIA station. The "intelligent device" differs from the self-aware one in that it can not only identify itself but it also implements the TSIA standard interface for its device class. The "ideal device" does all of this plus it employs the same data transport mechanism as the station's internal communications bus and, therefore, is able to attach directly to the bus without the need for an intervening proxy.

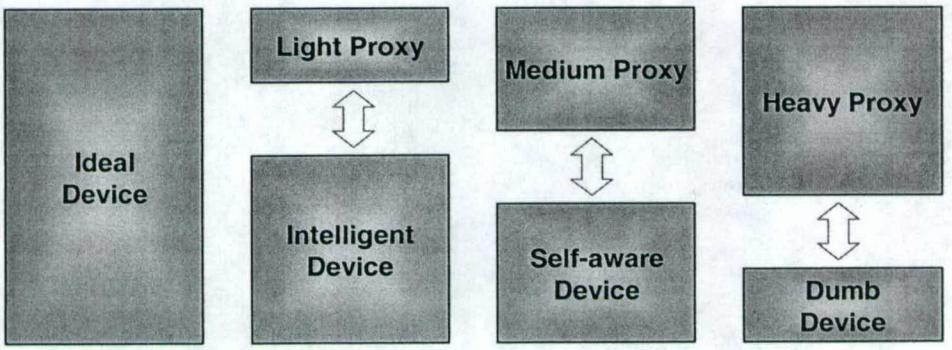


Figure 6. Devices and Proxies

Sixth, in TSIA, when an intelligent device or self-aware device is attached to the station, a "proxy factory" generates whatever proxy is required to adapt the device to the internal communications bus. If the station has never before seen this kind of device, the proxy factory mechanism will recognize this and will work with other station components to determine where on the Internet to look for the needed software and will then download and install this software automatically. In today's systems, integration of peripherals is done during engineering and production. No facility for end user integration of devices exists.

Seventh, protocol components in TSIA exist to deliver particular clinical functions, such as collection of images, to station users. When invoked, a protocols' first job is to determine what components are available to support its associated function. Each such protocol can have both preferred and back-up modes of operation. For example, in the case of image collection, the preferred mode of operation may be to route images to a remote station in real-time; however, the protocol may discover that the available image-generating device produces data at a rate that exceeds the capacity of the available

external communications channels. In this case, the protocol may try to acquire storage space where imagery can be buffered while being trickled out over external communication lines.

Finally, because station components need not be local, the TSIA station-level architecture makes it entirely possible to create a station in which only the peripherals, the devices needed to connect them to the network, and user interface components are collocated with the station's user and everything is "out there somewhere" on the Net. This sort of "thin client" approach to telemedicine station design has yet to appear in the telemedicine market place but will make it possible to deliver very inexpensive stations to end users.

How The Station Operates

As an illustration of how the station-level architecture is to work, consider the simple system shown in Figure 7. The ***user interface framework*** represents the top-level user interface constructs (menus, main buttons, etc.) used to initiate functions in the system. The ***pulse oximeter user interface module*** is a user interface component that is presented when a user needs to control a pulse oximeter and to display the data generated by the ***pulse oximeter***. In this scenario, the pulse oximeter accepts start and stop commands and outputs a continuous oxygen saturation waveform. The ***waveform statistical analyzer*** accepts waveform data and generates a variety of statistics that characterize the waveform (e.g., average, max, min). The ***pulse oximeter operation button (and event handler)*** represents that user interface control contained in the user interface framework that is used to start and stop the operation of the ***pulse oximeter protocol***. This protocol contains instructions regarding the kinds of components that are needed to support the protocol's operation, the ways in which these components need to be interconnected, and events that are to be monitored during the time that the protocol is active. The ***registry*** exists to allow components to discover each other's existence. It is assumed that the station knows about these hardware devices and that of all of the software components needed for successful operation are already installed on the station.

To start, when the station first powers up, the registry will scan its local environment for available resources. This will confirm the presence of the user interface elements, pulse oximeter protocol, and the station's context (not shown). Once this scan is complete, the registry will trigger the context that will examine its internal specification regarding desired station configuration and then interact with the registry to reserve needed resources. Once leased, these resources are directed by the context to interconnect in a particular way. In the case of this example, the result would be that the user interface framework shows a button for triggering pulse oximeter protocol and this protocol is associated with this button (i.e., the station knows that when that icon on the screen is pushed the pulse oximeter protocol is to be initiated).

Now assume that the station user adds the pulse oximeter to the station. This causes the device's presence to be registered with the station (Figure 7). Then, when the user selects the pulse oximeter operation from the user interface framework (Figure 8), the associated event handler instructs the protocol to "prepare" itself by leasing needed resources from

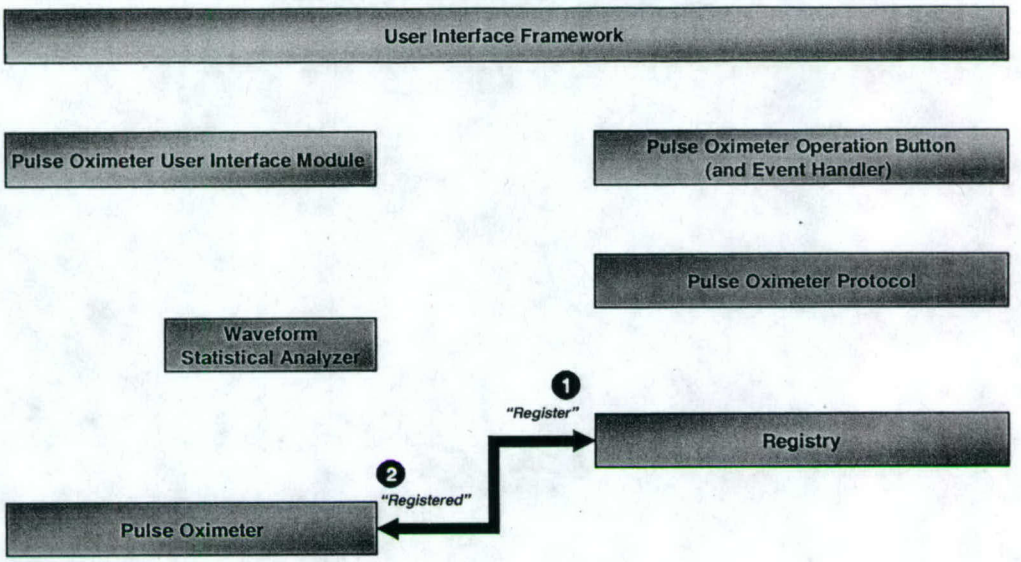


Figure 7. Registering Components

the registry and instructing each of the leased components to subscribe to specific services offered by other leased components (Figure 9). Once done, the protocol notifies the event handler which instructs the UI framework to display the pulse oximeter user interface module for the station's operator (Figure 10).

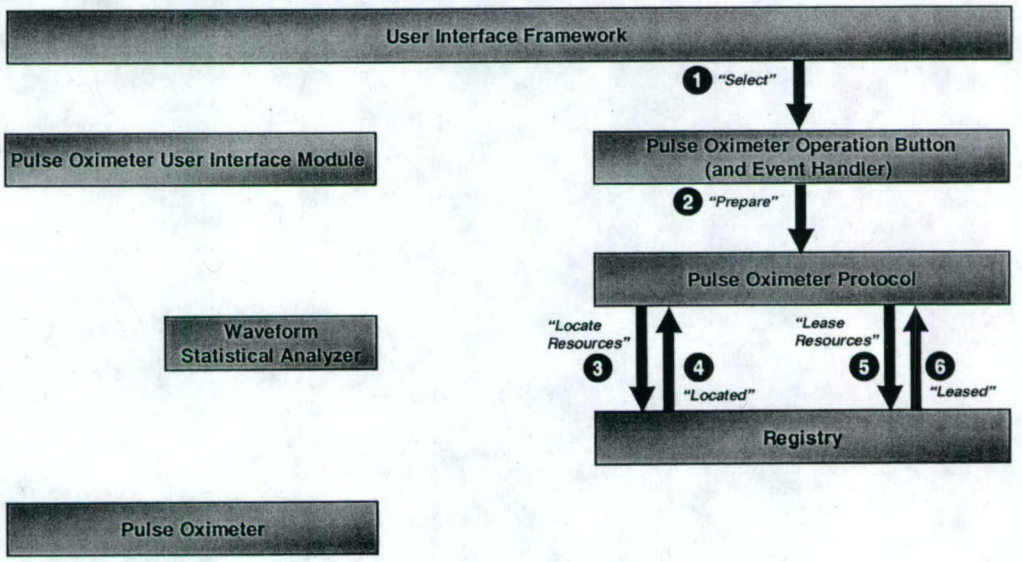


Figure 8. Initiating a Protocol

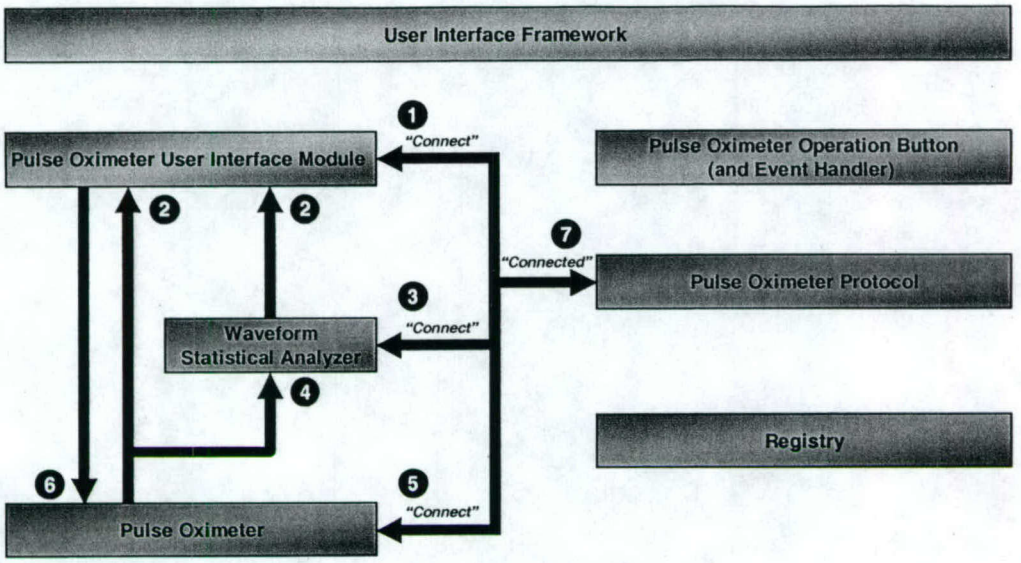


Figure 9. Connecting Components

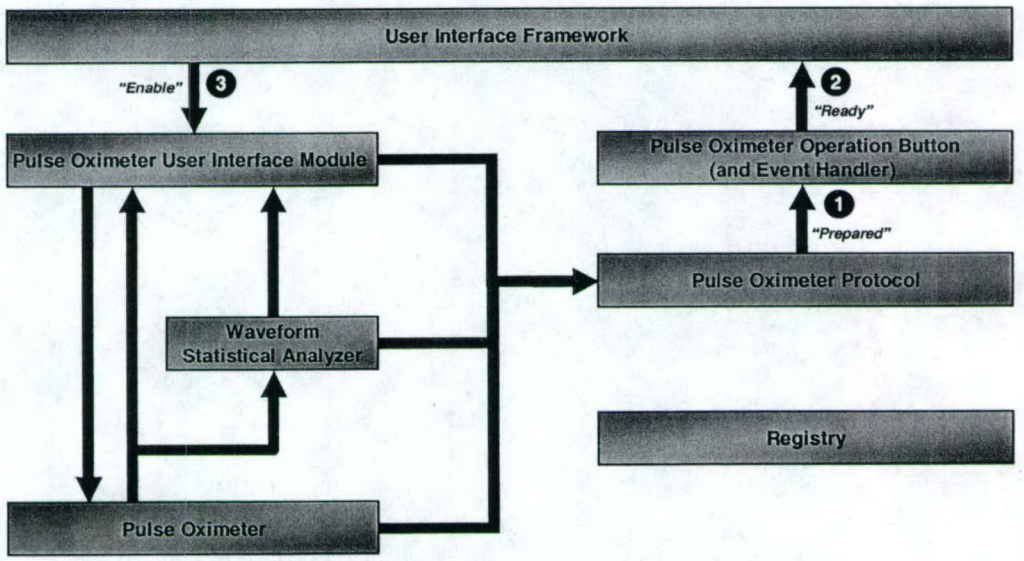


Figure 10. Enabling User Operation

System operation now proceeds with the user instructing the station to start (or to stop) taking pulse oximeter readings. While it is operating, the pulse oximeter sends its waveform data to its subscribing components (i.e., the pulse oximeter user interface module and the waveform statistical analyzer) and the analyzer sends the data on to its subscriber (Figure 11). If, during the operation of this network of components, any of these components experience an event that compromises its ability to support the protocol (e.g., the pulse oximeter is removed from the station), then the affected components notify the protocol which must then decide how to handle this situation.

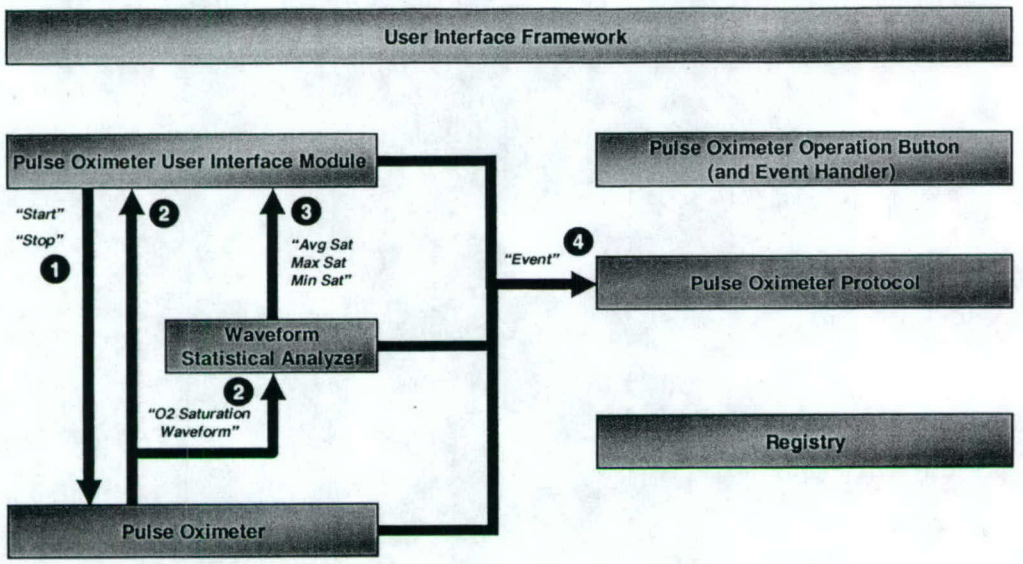


Figure 11. Using the Component

When finished with the pulse oximeter, the user may "deselect" the device on the user interface framework, which results in the pulse oximeter user interface module being disabled and the pulse oximeter protocol being told to terminate itself (Figure 12). In turn, the protocol instructs each of the leased components to terminate its connections. The protocol then notifies the registry that it is vacating its lease on these components and tells the user interface event handler that it is ending (Figure 13). The event handler then passes this fact on to the user interface framework that returns its interface to the same state that it was in before the pulse oximeter was first selected.

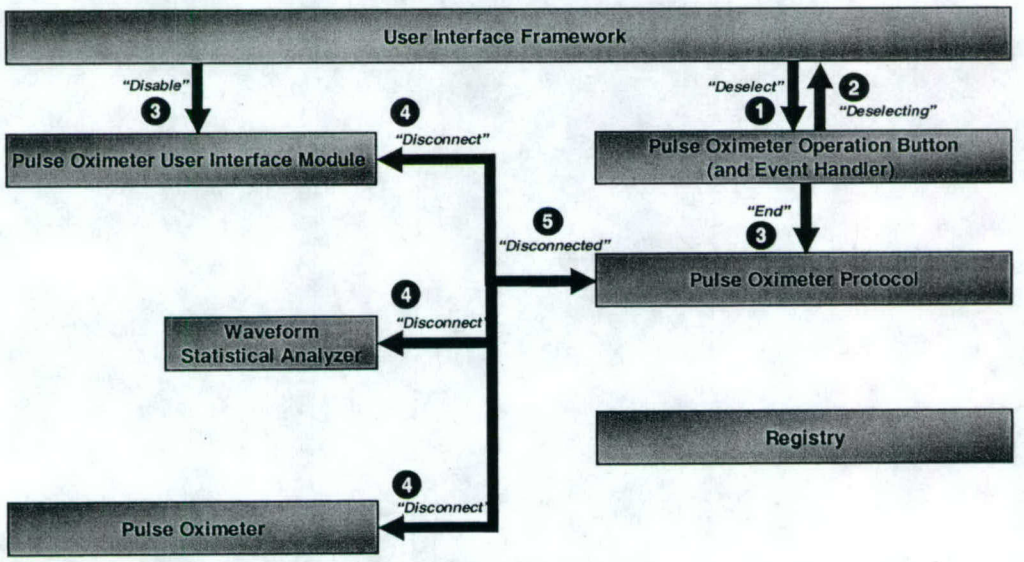


Figure 12. Disabling User Operation and Disconnecting Components

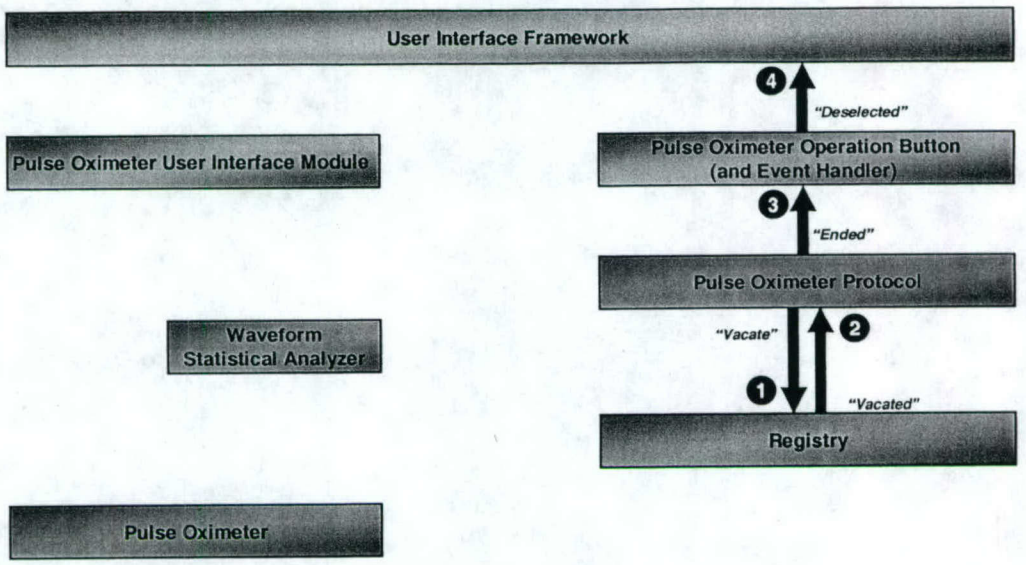


Figure 13. Vacating the Leases

Had the station not known about pulse oximeter before the device was added, the attachment of the device to the station would have triggered an additional process of identifying the new device and then retrieving and installing the infrastructure needed by this device. Exactly how this is achieved is discussed in the next section on the TSIA's network-level architecture.

The Network-Level Architecture

The goal of the network-level architecture is to allow independently designed and implemented stations to locate each other, explore each other's capabilities (subject to each station's access control rules), to negotiate with each other and with the networks that they will use to determine how a given session will be run (e.g., what quality of service requirements will be levied and what resources will be leased from each other), and to then conduct collaborative operations. Key architectural components (Figure 14) at this level include ***telemedicine stations***, ***registry servers***, and ***networks***. Stations are the collections of resources described in the first part of this section. Registry servers are computer-based systems that allow stations to advertise their location and, if desired, services and that allow stations to locate either specific stations or stations that offer particular kinds of services. Networks consist of those telecommunication components that carry information back and forth between stations in a telemedicine system. Depending on the nature of the security services ultimately employed in the TSIA, an additional component – ***security servers*** – may be required in order to allow secure interoperation between stations that have never before been connected to one another.

At this level, stations present two interfaces to other stations. The ***station registry interface*** provides a way for one station to explore the capabilities that another station can provide. These capabilities can include any of the resources discussed in the station-level architecture (medical devices, user interface devices, patient records, processing components, and even protocols and communications). The ***session management***

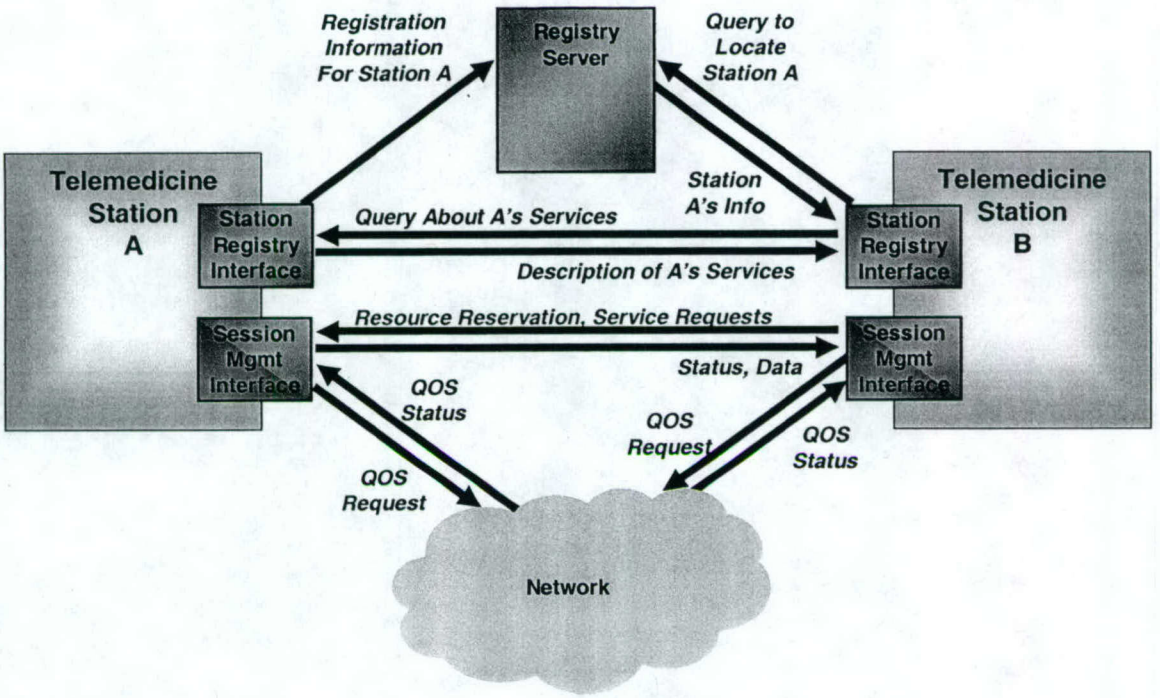


Figure 14. Network-Level Interoperability Architecture

interface provides a means for stations to lease these services for a period of time and for the quality of service components in a network to negotiate a contract with a station and to report, as needed, of the status of that contract.

What's Different About This

The TSIA station-to-station interoperation approach differs from current designs in several key ways. First, in traditional designs, each station capable of initiating contact with another station will store its own "address book" that allows a user to "call" these other stations. In contrast to this, TSIA's registry server approach that enables stations to locate other stations either by specific ID (e.g., "Find Mrs. Smith's station.") or by station attributes ("Find a server that can perform these kinds of analyses."). This White Pages™ / Yellow Pages™ approach to contacting other stations will be critical in a world where any station is meant to talk with any other station and stations are treated as much as communication systems as they are as medical devices.

Second, in most current telemedicine system designs, stations operating with one another will know in advance of contacting each another exactly what capabilities reside on each other's station. In the TSIA architectural approach, interoperating stations have no advance knowledge of each other's capabilities before contact is first made. This must be so if end users are to be allowed to customize their own stations on the fly. To compensate for this lack of a priori knowledge, stations in a TSIA system will begin their interactions by first querying each other to determine what capabilities are present.

Third, the TSIA approach incorporates session management into its station-to-station design so that stations can negotiate details of interoperation at a very fine-grained level. Included in this can be details such as specific transport protocol to be used for a given operation, desired quality of service parameters, and security mechanisms to be used.

Finally, as noted earlier, both a station using a remote station's capabilities and the station being used have the ability to self-configure as required. Just a station can download needed software components when a new device is added to the station, a station desiring to exploit remote station capabilities that it has never before used has the ability to automatically download and install software needed to support these new kinds of interactions. This downloading is not limited to things having to do with major components (e.g., patient records or medical instruments) but also addresses things like components that implement certain protocols or security mechanisms. Again, this sort of capability is central to delivering any station to any station interoperation.

How The Network Operates

The TSIA network-level architecture is built around six types of service:

- station registration,
- station location,
- station exploration and leasing,
- session management,
- session operation, and
- self-configuration.

Figure 14 depicts all but the last of these services.

When a station is started up for the very first time, it makes contact with a registry server and passes to this server information needed by other stations to identify it. In the figure, Station A is shown taking this step.

Once a station is registered, other stations querying the registry servers can determine how to contact the station (as is being done by Station B in the figure). Once contact is made, one of the stations will transmit queries regarding the other's capabilities and, in response, will receive back a description of what services the responding station can offer. Depending on security considerations, the list of available services returned may represent all available services or just a subset of the station's actual capabilities.

At this point, the querying station may generate a request to lease one or more of the services provided by the other station. For each such service negotiated, details related to how the associated session will be managed (e.g., how much bandwidth will be used, what protocols will be employed) are established. Once a lease has been granted and session management details negotiated, the stations are free to interact (e.g., carry out videoconferences or monitor and control remote devices). During this time frame, it is possible for the networks supporting the session to alert the stations when quality of service agreements can no longer be honored. When this occurs, the stations may renegotiate session management details before continuing operation. Similarly, if new services are requested before existing sessions terminate, the stations may need to

renegotiate the details (e.g. how much bandwidth is being used) of existing sessions in order to free up the resources needed to accommodate the new service requests.

Figure 15 depicts TSIA's approach to the last of the network-level services – self-configuration. This operation is initiated whenever a station is confronted with a component that it needs to employ but has never before seen. Using the device's unique identifier, the station is able to determine where a server containing software components needed to employ the device can be found (two possible techniques for doing this are to store a network address for the server within the device itself or to maintain a list of software servers in the registry servers used by telemedicine networks). Using this information, the station's installation manager contacts the software server asking that the needed software be returned. The package received by the telemedicine station is unbundled by the installation manager and the software components are installed and registered on the using station.

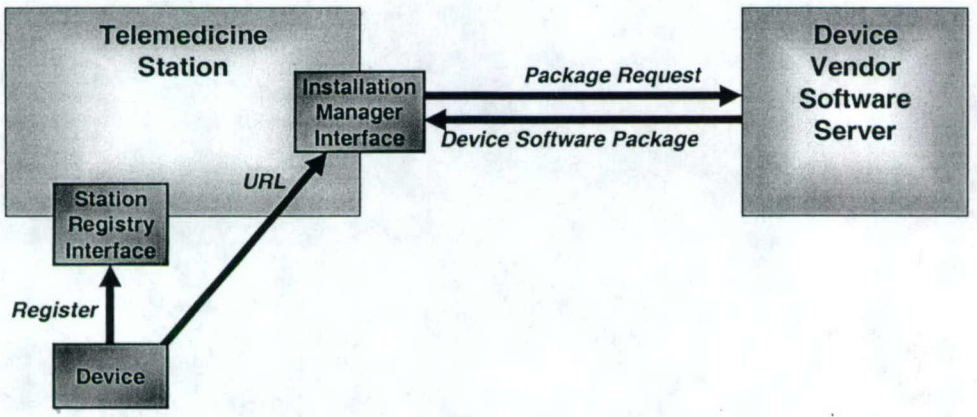


Figure 15. Self-Configuration Architecture

How This Approach Addresses Current Needs

Before leaving this discussion on the architecture, it is worth considering exactly how the proposed architecture addresses the various factors that shaped its development.

The architecture address the end user need for vendor independence by allowing stations from different vendors to interoperate and by allowing end users to mix and match station components at will to create systems that meet their own unique operational needs. Likewise, by defining an architecture rather than just some station-to-station protocols, this approach allows vendors to decide exactly where they want to compete. If they want to continue to deliver everything, that's great. If the want to be compete only in certain niches – be the best of breed in a given kind of component, deliver the best price-performance point for a given device type, etc. – that works as well. Also, nothing in this approach requires that stations implement all interfaces specified in the architecture. One vendor may choose to continue to employ proprietary, monolithic station designs but choose to implement some or all of the station-to-station features of the architecture so as to allow interoperation with other vendors' products. Another may continue selling systems that employ proprietary station-to-station protocols but choose to employ the

station-to-device elements of the architecture in order to minimize the amount of work required to add new peripherals to its product line. Allowing vendors to pick and choose in this way was intentional. The goal was give vendors complete control over their business strategy by allowing each vendor to decide what portions of the interoperability specification it would adopt and when.

The ability to define where in the architecture vendors want to make a living lowers the cost of entry into the telemedicine marketplace. With the proposed interoperability approach, a vendor no longer needs to build entire systems in order to build a business in telemedicine. This should foster the development of a large component vendor community, much as occurred in the PC world as a result of IBM opening up the details of its original PC architecture. If this occurs, vendors will naturally seek to fill gaps in capability. As a result, telemedicine systems will be able to address a wider range of clinical functions than is possible today. In turn, this should improve how clinically relevant mainline healthcare providers deem telemedicine-based approaches to be.

The goal of being able to integrate telemedicine systems with other clinical and non-clinical assets is accomplished in the architecture through the judicious selection of interoperability specifications from these other domains. As will be discussed in the final section of this paper, standards already exist or are under development that address many of the services, such as the exchange of patient records, identified in the Telemedicine System Interoperability Architecture. Adopting these specifications rather than reinventing new ones makes the desired integration possible.

The proposed architectural approach should help lower equipment costs both by reducing engineering costs, by reducing material costs, and by providing the necessary foundation for the creation of a consumer-oriented telemedicine market versus one that exists on government grants.

Making Technical Interoperability Real

During the last decade, the telemedicine community has been quite successful at identifying barriers to telemedicine's success and at working to remove these barriers. Recent changes in reimbursement policy are one example of this. If telemedicine is to ever reach its full potential, the same thing must happen in the area of technical interoperability. The follow sections propose specific objectives for the telemedicine community adopt and suggest steps that might be taken to reach these objectives.

Specific Recommendations

Agree on an Architectural Approach

At its heart, interoperability is about agreement. Given this, it is critically important that the telemedicine community reach consensus regarding what approach will be used to achieve technical interoperability. As noted earlier, the Telemedicine System Interoperability Architecture was never intended to be "the answer" to this question; rather, it is meant to be a starting point for discussion within the community. In the end, it matters very little whether TSIA is embraced wholly, in part, or not at all. What does

matter is that *something* is embraced and that this something has the wherewithal to carry telemedicine to wherever it ultimately needs to be.

Having said this, it is worth noting that not every "standard" in use today came into being via consensus. Microsoft Windows™ is a good example of a case where the market dominance of single vendor set the pace for the rest of an industry. While no vendors like this exist today in the telemedicine industry, it is entirely conceivable that a single large company with vision, financial wherewithal, and willingness to stay engage long enough to cultivate a marketplace could revolutionize telemedicine system design and become the de facto standard against which all other vendors' products are measured. As a first step, the telemedicine community needs to decide whether it wants to sit back and wait for such an event to occur or whether it wants to actively shape its own future.

Identify Which Interface Standards Will Be Used

Although the author cannot point to sources, he has heard the statement "the telemedicine industry doesn't want to be in the business of setting standards" made on several occasions over the last couple of years. To this, the author would say, "Yes, you're right and, no, you're wrong." In practice, there are a number of existing standards that could be used to address key parts of the proposed architecture. In some cases, these standards are quite extensive and have large constituencies supporting them. To propose anything other than embracing these standards would be foolish. To the degree possible, the telemedicine should avoid reinventing the wheel.

At the same time, there are certain parts of a telemedicine system that are distinct from other domains. As a result, there are certain portions of the architecture for which no good standard is available off-the-shelf. If the telemedicine community decides to pursue technical interoperability, then in these telemedicine-unique areas, the community will have no choice but to accept the task of developing these missing standards This is not to say that the telemedicine community need do this unilaterally. In some cases, there will be other standards development organizations that will have an interest in these new specifications. Developing partnerships with these organizations where appropriate will be of use both in the near term as specifications are being developed and in the long term as these standards need to be maintained.

Promote Rich Markets in Plug-and-Play Devices and Other Station Components

An architecture and standards by themselves will be meaningless if steps are not taken to transform the telemedicine peripherals industry. Telemedicine systems built to date have made use of medical instruments designed for use in non-telemedicine settings. As technical interoperability standards are being forged, steps must be taken to promote the development of a rich array of peripherals that implement these standards and that are designed from the start with telemedicine operating environments in mind. They new devices will be meant for use in non-standard places, meant to be employed by a broader range of users, etc. (Figure 16). Those attributes shown in italics in the figure are of major importance in realizing this objective.

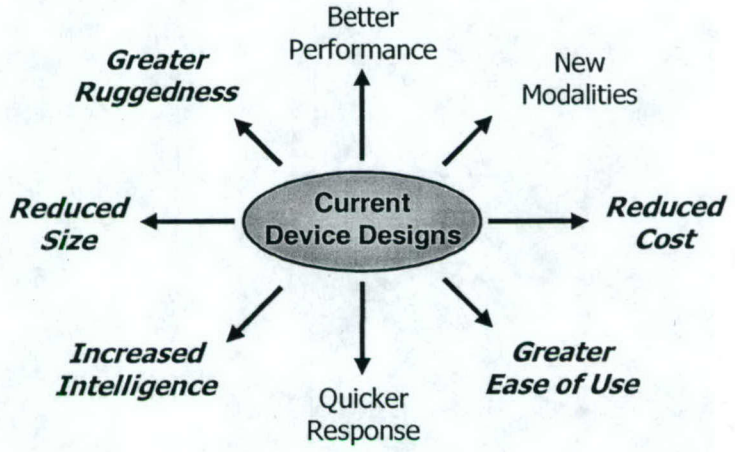


Figure 16. Directions of Improvement In Device Design

Address Device Certification

Given that flexibility is a central theme of an architecture like TSIA, one of the concerns that must be addressed by the architecture is the safety of stations that are composed on the fly by end users. Steps will be need to be taken to ensure that the architecture has an appropriate assurance approach and that this approach is worked out in conjunction with organizations like the FDA who are charged with ensuring the safety of such systems. One of the worst things that could happen in this area is for these authorities to dictate that stations that are to be fielded must be certified as a whole and that the addition of a new component to a certified system necessitates a whole new certification process.

Reexamine the Role of Telemedicine in Healthcare

Finally, TSIA is predicated on the belief that the architectural approaches used in typical telemedicine systems today do not have the power to do much more than service the operational concepts already being pursued by the telemedicine community and that there is more that can be done with computer-mediated approaches to care delivery. The question is exactly how much more.

Ringel and Bauer's assertion that "telemedicine will ultimately revolutionize healthcare – restructuring virtually every relationship and activity that define late twentieth century medicine" [13] seems to this author to be a statement of the obvious; however, it is not clear where in the healthcare community serious thought is being given to what this kind of revolution might entail. All of today's leading healthcare improvement initiatives seem to be evolutionary in nature – focused largely on making incremental quality improvements in existing care delivery structures.

Given its mix of people who understand the realities of the clinical world and people who understand the transforming potential of emerging computing and communications technologies, the telemedicine community is ideally suited to lead a revolution in care delivery; however, this will only occur if the community takes deliberate steps in this

direction. It must stop viewing itself as existing solely to address needs of fringe populations. It must see its mission as not just research into esoteric delivery methods but as reformation of mainstream care delivery processes. It must strive to foster the development of a robust, self-sustaining telemedicine system vendor community and must work to open up markets for this community to service.

A Path To The Goal

So what practical steps might be taken in the near term to move the community in this direction? What will be gained if these steps are taken?

Establish a Working Group on Architecture

As a starting point, a telemedicine architecture working group should be established. This body's primary goal will be to develop consensus within a critical mass of players in telemedicine regarding the technical interoperability approach that the community wants to pursue. Once this task is complete, the working group should serve as an architectural review board for more detailed products developed by others working on specific aspects of the selected architecture interoperability approach.

Develop Groups Around Different Components and Interfaces

If an approach to interoperability is adopted, it is likely that details of various portions of the approach will need to be worked out. In support of this, groups should be established around different portions of this approach (e.g., networking, medical devices, medical records) whose job is to bring these portions of the overall approach to completion. Inasmuch as existing standards work addresses many aspects of what telemedicine systems require, these groups should establish relationships with the appropriate standards development activities with a view to shaping/extending these standards in ways that address the needs of the telemedicine community.

Create Reference Implementations of All Interface Specifications

To speed adoption of the community's interoperability approach, software that implements each its *interfaces* should be developed and made available in open source form. Engaging universities to do this work and making the resulting source code available through a mechanism such as Source Forge [14] would be a reasonable way to accomplish this.

Underwrite a Demonstration of the Interoperability Approach

A multi-vendor demonstration should be underwritten that addresses both station-to-station operations and station-to-device operations. The goal should be to show that different vendors can interact with each other's stations and that a suite of new medical instruments developed explicitly for use in telemedicine environments can be swapped between different vendors platforms. It is important to note that the thrust of this work is engineering and not advancing the science of medicine. As such, these demonstrations should not be slaved to clinical studies but should stand on their own as engineering development activities and proof-of-principle demonstrations.

Establish an Initiative on the Future Telemedicine-Based Care

A multidisciplinary study should be pursued with a view to conducting a "clean sheet" rethinking of the role of telemedicine in the future of care delivery. The tone of this study should be possibilistic in nature (i.e., should not be constrained to examining only incremental changes in current practice). A primary goal of this effort should be to identify principles that govern the applicability of over the computer-mediated approaches to healthcare.

Conclusions

Telemedicine has the potential to revolutionize healthcare delivery but doing so will require that certain engineering-related issues, such as the lack of a community-accepted technical interoperability approach and the need for a new generation of devices designed for use in telemedicine, be addressed along the way. With respect to this first issue, adopting an architecture-based approach to interoperability will carry the telemedicine community farther than approaches based on a simple enumeration of station-to-station interoperability protocols. Such an approach would not be (should not be) a "clean sheet" exercise. A substantial body of relevant work already exists in various standards development organizations and, wherever possible, the telemedicine community show exploit whatever has already been done to the extent that this work meets the community's needs. Whether the specific architecture presented in this document is embraced is unimportant. What matters is that the telemedicine community commit to pursuing some approach to telemedicine and it is important that this be done sooner rather than later. If the telemedicine community chooses to simply march in place, it will very likely find that trends in technology and shifting demands in the world of healthcare will force the kind of solution proposed in this paper to emerge from some other quarter. The decision to be made then is not whether or not interoperability will ever be addressed but whether or not the telemedicine community wants to be the one that addresses it.

References

[1] Health Level 7 (HL7) Standard, http://www.hl7.org/Library/standards.cfm

[2] SNOMED International, http://www.snomed.org/

[3] Buxton, Peter J. and D. M. Lam, "NATO Telemedicine Interoperability Study (TiOPS)", presentation to the American Telemedicine Association 2003 Annual Meeting, March 2003, Orlando, FL., http://www.americantelemed.org/news/2003_presentations/M3e3.lam.htm

[4] Ringel, Jeffrey C and M. A. Bauer, Telemedicine and the Reinvention of Healthcare, McGraw-Hill Professional, 1999

[5] Ross, Muriel D., "Medicine in long duration space exploration: the role of virtual reality and broad bandwidth telecommunications networks", *Acta Astronautica* Volume: 49, Issue: 3-10, August - November, 2001, pp. 441-445

[6] National Cable Television Association, http://www.ncta.com

[7] WebOptimization.Com, http://www.websiteoptimization.com/bw/0403/ (statistics taken from Nielsen / NetRatings)

[8] Foster, Ian, C.Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", http://www.globus.org/research/papers/anatomy.pdf

[9] Muschamp, Paul, "An Introduction to Web Services", BT Technology Journal, Volume 22, No. 1, January 2004, pp. 9-18

[10] Green, Rebecca P. and M. Landt, "Home Sodium Monitoring in Patients with Diabetes Insipidus", The Journal of Pediatrics, Vol. 141, No. 5, November 2002, pp. 618-624

[11] Weed, Lawrence L., "New Connections Between Medical Knowledge and Patient Care", BMJ 1997;315:231-235 (26 July)

[12] Craft, Richard L., The Telemedicine System Interoperability Architecture: Concept Description and Architecture Overview, http://telemedicine.sandia.gov

[13] Ringel, Jeffrey C and M. A. Bauer, Telemedicine and the Reinvention of Healthcare, McGraw-Hill Professional, 1999

[14] Source Forge, http://www.sourceforge.net

Semantic interoperability: bridging communication gaps between stakeholders

Haim Kilov
Independent Consultant, and Stevens Institute of Technology
haimk@acm.org

> "When a number of drawings are made after one pattern, though they may all miss it in some respects, yet they will all resemble it more than they resemble one another; the general character of the pattern will run through them all; the most singular and odd will be those which are most wide of it; and though very few will copy it exactly, yet the most accurate delineations will bear a greater resemblance to the most careless, than the careless ones will bear to one another."
>
> Adam Smith, *The Theory of Moral Sentiments*. 1759.

Introduction

In thinking about and discussing interoperability, we observe that the systems that have to interoperate need not be computer-based. Specifically, we may — and probably ought to — look at business-based and IT-based stakeholders as components of two or more systems that often have serious difficulties in communicating.

The proverbial communication gap between business and IT experts has led to substantial problems in interoperability (both syntactic and semantic) between different stakeholders of different (such as business and IT) systems, leading, in turn, to significant monetary losses together with loss of customers' trust and patience. The paper will demonstrate not only the problems but also the solution — a clear separation between the business and IT domains based on an explicit usage of a system of concepts common to all domains and understood by all stakeholders. These elegant concepts come from exact philosophy[1], mathematics, programming and systems thinking. They have been successfully used not only in theory but also in industrial practice (including international standards such as the Reference Model of Open Distributed Processing — RM-ODP), and in teaching of business and IT modeling. The paper shows how a system of exactified concepts and approaches, especially such concepts as system, type,

---

[1]Our interest in and usage of exact philosophy may be explained, for example, by Bunge's observation that concepts and hypotheses are philosophical because they occur in a large number of fields of inquiry [B2004]. This was recognized in the curricula of (at least) the early universities. Of course, business analysts as generalists who work on analyzing any systems, and (information) system designers belong to the esteemed class of people who use and sometimes develop such concepts and hypotheses.

relationship, composition, pattern, name in context, etc., has been used to understand and specify the semantics of non-trivial industrial business and IT systems, thus establishing a basis for successful communication between business and IT experts, that is, for semantic (and sometimes syntactic) interoperability.

Conceptual requirements for interoperability

Let us start with a recent description of two familiar kinds of interoperability. "Syntactic interoperability is all about parsing data correctly. Semantic interoperability requires mapping between terms, which in turn requires content analysis. This requires formal and explicit specifications of domain models, which define the terms used and their relationships. Such formal domain models are sometimes called ontologies." [SW2004] This description may be used not only for dealing with computer-based information systems but also — and perhaps more importantly — for dealing with *human stakeholders communicating with other humans or with computer-based systems*.

Of course, this reference to domain models and relationships is not new. Walter Bagehot, one of the founders of modern money markets, suggested in 1873 the same approach in order to understand and communicate about the objects in the money world: "[t]he objects which you see in Lombard Street, and in that money world which is grouped about it, are the Bank of England, the Private Banks, the Joint Stock Banks, and the bill brokers. But before describing each of these separately we must look at what all have in common, and at the relation of each to the others." [B1873].

Recognition of *importance of domain models* has not been always forthcoming in the context of computer-based information systems. However, in any engineering discipline, domain understanding and specification is essential before requirements can be understood and formulated, and of course, a system can be developed only on the basis of well-understood requirements. This was emphasized, for example, by Dines Bjørner: "domain, requirements and software design are three main phases of software development". As we all know, sometimes IT system requirements exist only "in the collective heads" of the developers, leading to more or less serious failures described, for example, in Peter G. Neumann's "Risks to the Public" column in *Software Engineering Notes*. Therefore interoperability between stakeholders during domain modeling, as well as during requirements understanding and specification, is essential for success in development of any systems, including software systems.

The communication gap between business and IT experts can be exactified as the absence of interoperability. Firstly, often there is no *syntactic* interoperability because business stakeholders are often unable to parse data used by IT stakeholders. Of course, business experts cannot (and should not!) read code, but they also often cannot read specifications written by IT experts using notations[2] (or terminology) which are overly complicated or alien to business. Here is an example from the OOPSLA'99 keynote (on e-business) by Stu Feldman: "We need to understand the

[2]Business experts have no time or desire to study — for 5 days, 8 hours a day — the many facilities of a notation using toy examples, as often happens when popular "powerful" notations are taught.

domain before addressing software. … Business models are the basis of an organization's entire activity. They are to be understood by CEO and CFO, not just by CIO; and therefore explained without 'method calls will have an XML representation'."

Secondly, often there is no *semantic* interoperability between business and IT experts and also between different business experts, as well as between different IT experts. This happens because the same data (such as "meaningful names") may be and often are interpreted differently by different stakeholders, and in the absence of an explicit domain model these differences may not be discovered until it is too late. Here is a somewhat familiar example attributed to James Schlesinger and quoted in the *Forbes* magazine: "...when the Marines are ordered to '*secure a building*', they form a landing party and assault it. The same instructions will lead the Army to occupy the building with a troop of infantry, and the Navy will characteristically respond by sending a yeoman to assure that the building lights are turned out. When the Air Force acts on these instructions, what results is a 'three year lease with option to purchase'…".

The problems of semantic heterogeneity have been observed in software engineering for quite a while. Relying on syntactic interoperability — in particular, on "meaningful names" — leads to failures. Grace Hopper observed in 1957 (!) that "[w]hile the computation of the square root of a floating decimal number remained the same in Pittsburgh, Los Angeles, and New York, the computation of gross-to-net pay obviously did not remain the same even in two installations in the same city" [H1957]. The same kind of problems still exists today, and we still often encounter statements like "everyone knows what a 'patient' is", or "everyone knows what a 'trade confirmation' is". Various tools — modern and not so modern — usually avoid these issues, and for a good reason: understanding and solving semantic interoperability problems requires human intervention that cannot be replaced with any tools.

Clearly, problems described above are not specific to software systems. They have been encountered in any kinds of systems, independently of their realization mechanisms.

Obstacles for interoperability in general and computer-based systems

Interoperability problems have been acknowledged by many business and IT experts. Solution attempts have existed for quite a while, but often they resulted in not much more than warm and fuzzy feelings during meetings. In many if not most situations, failed attempts were based on various more or less fashionable information technology tools and methodologies instead of clear and explicit domain ontologies. On the one hand, most specifications have relied on (a lot of) tacit assumptions which are clearly different for different specification readers. On the other hand, even explicit fragments of specifications presented using box-and-line diagrams (as all too often various architectures and even business plans have been shown), or in natural language, lead to serious problems because different people will interpret the specification differently. Indeed, in order to transmit a message from one person to another without loss of meaning, the author and recipient(s) of the message have to use the same ontology and the same notation. Using the same natural (colloquial) language as a notation is not sufficient since in order to preserve meaning we need also the same context, the same language experience, language norms, cultural tradition, and

so on [L1990], and these properties of different people are often implicit and (very) different. At the same time, a restrictive artificial language *with precisely defined semantics* that does not have contexts, cultural traditions, and so on, can guarantee an adequate transmission of a message's *semantics*, provided, of course, that it can be adequately represented in that language. As a somewhat crude approximation of such a restrictive language, we may consider "legalese" in which, for example, laws (providing "the same context") and contracts are written.

Let us generalize these observations and consider two aspects of interoperability problems in general systems. On the one hand, different stakeholders may (and often do) "live in different worlds", so the things, relationships and actions of these worlds (like those of "business" and "IT") may *seem* to be very different and even incompatible. On the other hand, as noted by Grace Hopper, even in the same world the same names may (and often do) refer to quite different things. In other words, there appears to be a very substantial communications gap — different stakeholders just do not use semantically the same language.

When communicating humans encounter semantic interoperability problems they can (and often do) resolve these problems in an informal manner by asking questions, providing definitions, examples and counterexamples, etc. In doing so, tacit or (somewhat) explicit domain models used by the humans are elucidated (to a certain extent) and compared. Clearly, recognizing that problems exist is an essential prerequisite for solving them. This recognition — also often informal — may or may not happen, or may happen partially. However, when (some) communicating agents are computer-based information systems such informal problem recognition and resolution is impossible because, as Dijkstra observed a while ago, computers execute programs as written and not as dreamt.

More recently, the recognition of importance of ontologies (see, for example, an overview in [WW2004]) could have become a much-needed innovation — not even a radical one — needed to solve the interoperability problems..

However, ontology development today is in a poor state. In many cases, it has been replaced with an emphasis on (a large number of) logic and ontology languages, in the same manner as programming has been replaced with an emphasis on various "baroque programming languages" blurring our vision "by the wealth of their mutually conflicting 'powerful features' " (Dijkstra), or in the same manner as analysis has been replaced with an emphasis on "Undefined modeling languages" (Parnas). As a (or the) result, the complexity of the domain and problems has been replaced with the complexity of the language (as Dijkstra observed, these languages were used to express in a funny way the usually given algorithms). Furthermore, in the context of modeling and ontology languages, methodologies are being created for the sake of understanding how to use the usually complex tools. Because of the lack of any generally accepted processes and methodologies, the tools exist independently and have little support for or concern with interoperability (Ken Baclawski). In addition, most conceptual modeling activities have proceeded without the benefit of theory [WW2004]. These often buzzword-compliant approaches are tinkering (Bunge) rather than engineering ones.

The primary goal of a programming language is accurate communication among humans. Clearly,

the same is true about a modeling (or ontology) language. With many currently popular languages, this goal has not been reached.

As a result, even in cases when a specification exists and has been agreed upon by the relevant stakeholders, it may not guarantee interoperability of compliant systems, computer-based or otherwise. For example, if one vendor of a supposedly compliant product interprets the specification one way and another vendor interprets it another way, then both will claim compliance yet the products won't interoperate, and nobody can say for certain which interpretation is "the right one". For another example, different business experts may agree on the apparent validity of the business specification, but may interpret tacit underlying assumptions in different ways, and therefore their understandings of the specification — composed of the explicit parts and the tacit assumptions — will differ. If these understandings are substantially different then informal discussions may result in clarification, but if the differences are less substantial then (IT and) business problems may and probably will result. For a lot of examples, see Peter G. Neumann's "Risks to the Public".

A system of common concepts: The way to address obstacles to interoperability

The need for effective patterns of reasoning. Information management systems are becoming more complex and non-trivial since they have to serve the needs of complex, non-trivial and rapidly changing businesses. In order to succeed in understanding, specifying, designing and developing information systems we should do better than use rigid methodologies and step-by-step approaches.

This is difficult and perhaps unusual. As E.W.Dijkstra observed a while ago, "many students don't want to be shown effective patterns of reasoning, they want to be told what to do. ... They expect a so-called 'complete methodology'... and complain when they don't get what only the quack can provide. (We just addressed a bunch of industrial computing scientists, and the above phenomenon was alarmingly pronounced.)" [D1986]. Dijkstra's observation might have been dismissed by some as a useless academic exercise; however, more recently we read in the very pragmatic "ComputerWorld" (an article by Paul Clermont, June 7, 1999): "IT strategy today needs less methodology. Instead, it needs more content knowledge and creative thinking. ... We do need frameworks and principles, but specific methodologies ... must be tailored to the business situation and culture or even invented on the spot. Only well-seasoned consultants have the necessary sensitivity, seniority, flexibility and creativity to make this work. If a consulting firm emphasizes proprietary methodology and computer-based tools more than its experienced and insightful people, beware!". These warnings are even more appropriate now, after the "dot-com" epoch when people "suddenly realized that they had invested a fortune based on a few beautiful graphics that were laughably called a business plan" [L2001].

Instead of using "junk food" — the metaphor for simplistic methodologies we owe to Paul Clermont — we should better use effective patterns of reasoning that help at all stages of information management lifecycle, as well as in business management, independently of any possible

computer-based realization of its fragments.

Where do we get these patterns of reasoning? We certainly do not *want* to invent them for each and every project or stage of a project. Fortunately, we do not *need* to do that either: an excellent and well-structured system of common concepts on which patterns of reasoning may be based, already exists. It was defined in an abstract, precise and concise — elegant! — manner as an international standard: the *Reference Model of Open Distributed Processing (RM-ODP)* standardized by the International Standards Organization (ISO) in 1995.

RM-ODP specifies semantics — meaning — in a manner that is syntax-, methodology-, and tool-neutral. It provides *precision without programming*[3]. The Foundations of RM-ODP are very short — only 18 pages. Every concept there is precisely defined in clearly structured English within the context of other precisely defined concepts. In other words, the reader does not have to figure out what a particular term means, and neither does the reader have to rely on tacit assumptions left undefined since "everyone knows what this means" (but do they know and mean the same things?). How often have we heard and perhaps participated in violent discussions about the meaning of a class, a type, a component, a composition, conformance, or, more recently, of a "business object"? The relationships between these concepts also are precisely defined in RM-ODP.

RM-ODP eliminates much of the work that would otherwise be required by each organization to develop its own similar but proprietary guidelines — patterns of reasoning used to understand and to specify businesses as well as to specify, design and develop information systems for these businesses. RM-ODP (hopefully) will also provide the incentive to many business and IT organizations to follow similar approaches in developing specifications, leading to industry standards. RM-ODP has already been substantially used in creating other important standards, such as ISO standards — General Relationship Model and Trader, and OMG standards — UML Profile for Enterprise Distributed Object Computing, Model-Driven Architecture, and others, as well as industry-specific (vertical) business-specification standards.

Is this a radical novelty? There is nothing radically new here. The need to elucidate the definitions of things, actions, and especially the *structure (relationships)* of a system for understanding of that system has been noted by many authors, both in modern-day information management and systems thinking, and much earlier (recall the quote from Bagehot, for example). Moreover, the concepts essential to understand and specify the semantics of system components and structure have been formulated and discussed in IT, mathematics, philosophy, and system analysis for a while. RM-ODP and the international standard used to describe relationships in more detail — the General Relationship Model (GRM) — define a *system* of these concepts and are based on these ideas. This system includes such concepts as system, abstraction, viewpoint, level, object, action, state, behavior, type, template, composition, refinement, contract, name, context, invariant, pre- and post-condition, failure, error, and conformance. These *semantics*

---

[3]Some IT experts claim that "business people cannot understand precision" and that therefore business experts should be provided only with various narratives and pictures instead of precise specifications. These condescending claims are obviously wrong since business experts have understood precision and made precise decisions for millennia.

concepts are not associated with a particular technology approach (such as object orientation) and are neutral with respect to representational or tool-related issues.

Many of these concepts are quite familiar to good programmers and analysts. In particular, most have been around in programming since the mid-1960s. Many have also been successfully used in various modeling approaches within the framework of the three-schema database architecture. Some of them have been used in engineering and other areas of human endeavor (such as business and law) for centuries. The RM-ODP definitions are theoretically sound (based on mathematics, as demonstrated, for example, in the Architectural Semantics part of the standard), have been successfully used in practice and are independent of any particular representation or implementation mechanisms. In the same manner as businesses in the US rely on the standard Uniform Commercial Code, system specifiers ought to rely on the standard RM-ODP.

Is RM-ODP really useful? Here is an example of using RM-ODP to specify distributed systems (from the European Air Traffic Management System Architecture Workshop, held at EUROCONTROL headquarters in Brussels (Belgium) on June 11-13, 1996): "The application of RM-ODP provided insight into a number of interoperability issues ... The RM-ODP ... quickly showed the importance of correct focusing, scope, interpretation and representation within the descriptions. The RM-ODP approach impels the analyst to state clearly the focus and scope at the beginning of the process... During the work, the importance and feasibility of being able to check for completeness and self-consistency in requirements became clear through the ODP approach. In all the case studies, valuable insights into the systems... were gained and a number of problems with extant specifications highlighted. ...The model provides a common, consistent and incremental approach for describing the goals, objectives and behaviour of systems in detail. Hence, it offers support for life-cycle management and for strategy studies. In a similar way, it also offers support for COTS procurement – both for the procurement specification and for the suppliers' system specification and design specification."

Look at this evidence. Not only was RM-ODP useful at the specification stages; it was valuable at all stages of information management lifecycle including procurement and evaluation of suppliers' systems. *It is just good practice!*

There exists a lot more evidence of this kind, not only in air traffic control but also in less exotic areas of telecommunications, finance, insurance, document management, medical, pharmaceutical and other industries, as well as in management and strategic consulting. For example, RM-ODP (together with GRM) was used:

- to elucidate and describe various architectures in a large international financial institution,
- to provide a successful communication mechanism for stakeholders in business process reengineering projects (and to describe these projects from specification to realization),
- to specify COTS software components in a simple and understandable manner so that the semantics of these components became clear to their users,
- to create and use simple and elegant (and complete) business specifications of various financial domains (such as accounting, trading, exotic options, etc.) used by large financial firms in their work, both for information system creation and for business process change,

- to specify products by healthcare software vendors transforming creation of these specifications "from craft to a formal science",
- to develop a complete human resources model for DoD,
- to formulate a clear model of document management separating the concerns of content, logical design, and physical presentation,
- to provide a foundation for business decisions related to mergers and acquisitions,
- to elucidate and formulate fragments of the UML metamodel,
- to describe business strategies,
- to describe various business patterns,
- to develop ontologies of various business domains,
- to build the industry library in the pharmaceutical industry,

and so on. Some of these applications were described by satisfied clients in literature [KB2003 (several papers), EDOC2001, KMS1996, K2001a, HE1999, KA1997, G2001, SKB2001, LQS, TS1998, R1999].

Separation of concerns based on a common foundation. RM-ODP makes it possible (although not trivial!) to formulate a specification in a disciplined manner. Specifications will be read by people who are non-experts in specifications. This especially applies to *business specifications* all too often reduced to the elusive "business rules" (which are "in the code").

Discipline means precision and abstraction. *Precision* means (among other things) that a developer will not have to invent business rules that have not been described at all or have been described in an ambiguous or incomplete manner. *Abstraction* means (among other things) that a subject matter expert will not waste time and effort trying to understand business rules in terms of a particular computer-based implementation. Business rules (and a business enterprise in general) should be specified in terms of abstract and precise concepts understandable to any good subject matter expert, analyst, or developer — and yes, to a non-IT manager. (Recall the quote from Stu Feldman, above.) These concepts need not be reinvented for each project. They are the same for all kinds of specifications — thus providing *an excellent foundation for interoperability* — and have been formulated and described in a clean and methodology-, tool-, and representation-independent manner in RM-ODP and standards based on it, such as the General Relationship Model (GRM).

The basic concepts defined in RM-ODP and GRM can be — and have been — used not only to describe any kinds of traditional businesses, but also to describe the essentials of any existing or to-be-created information systems (computer-based or not). In this manner, the business and IT stakeholders are able to use a common set of concepts and therefore to communicate in a meaningful manner. Of course, the syntactic representations used by different stakeholders to represent the same semantics may differ, and also of course, different stakeholders may be interested in different levels and viewpoints when describing the same system, but the underlying semantic framework still remains the same.

While using the same system of concepts for all kinds of specifications, we should explicitly separate business from IT system specifications because business and IT ontologies are different.

(As a well-known example, a patient is not the same as one of patient's records.) Within each specification we should separate concerns as soon as we see that the specification (including a program — a specification for a computer system) becomes too complex for human understanding and is in danger of having "too much stuff". This is what abstraction is all about, and this is why abstraction is one of the first concepts defined in RM-ODP. ("Precise" is not the same as "detailed", and therefore being abstract does not mean being imprecise. Good specifiers, in the same manner as good engineers, postpone decisions and do not get drowned in details. The higher the level of abstraction the more important it is to be precise!)

Well-structured specifications: relationship semantics. A specification should be well-structured since only in this manner it can be carefully considered, read and understood. The structure of a system is "the collection of relations among its components or among these and items in its environments" [B2004]. Therefore precisely defined relationship semantics is essential for reading, understanding, and creating any specification. The semantics of all relationships is defined by means of their invariants referring to properties of relationship participants. And fortunately, it has been possible to specify the structure of a large number of diverse business and IT systems — such as financial derivatives, document management, UML metamodels, messaging, and system architectures — using only three kinds of generic relationships: *composition*, *subtyping* and *reference*. Semantic definitions of these relationships have been around for some time (see, for example, the exactifications in RM-ODP and GRM, as well as the information modeling text [KR1994]), and it is important and instructive to observe that these definitions substantially use *property determination*. In particular, the existence of emergent properties of a composite is the defining characteristic of the composition relationship (see [K2002, B2004] for many examples). Thus, it also becomes blindingly obvious why an often encountered statement "this [named] line between these two boxes formally represents the relationship between these two things" does not convey anything at all about the relationship semantics, and therefore why box-and-line diagrams are inadequate for understanding and decision making.

Abstraction levels and viewpoints. Precision (exactification) is not sufficient for human understanding. Indeed, hundreds or thousands of pages of precise material are (almost) useless if the material is not well-structured. In other words, understanding requires abstraction — "suppression of irrelevant detail" [RM-ODP], so that essential (for a specific level or for a specific viewpoint) aspects of a specification are clearly separated from accidental ones. Clearly, this is not a new approach, but it has not been stressed in most syntax- , methodology- or tool-oriented texts. And it is instructive that the concepts of abstraction, levels and viewpoints are among the first ones to be described in RM-ODP.

RM-ODP uses abstraction within the context of *levels*. In particular, it notes that "fixing a given level of abstraction may involve identifying which elements are atomic." More specifically, the concept of composition is defined using abstraction levels: "[a] combination of two or more [items] yielding a new [item], at a different level of abstraction. The characteristics of the new [item] are determined by the [items] being combined and by the way they are combined." This definition corresponds very well with the treatment of composition in philosophy [B2004] and elsewhere, including such classic as Adam Smith's *Wealth of Nations* (for details, see [K2002]).

RM-ODP also uses abstraction within the context of *viewpoints* — "form[s] of abstraction achieved using a selected set of architectural concepts and structuring rules, in order to focus on particular concerns within a system". All viewpoints are based on the same system of basic concepts. RM-ODP specifies five basic viewpoints — enterprise, information, computational, engineering, and technology. It is possible to define correspondences between viewpoints (RM-ODP shows how to do that and provides some examples), but often, one viewpoint cannot be defined in terms of another. Of course, the five basic viewpoints are not the only ones that may be used to describe a system. In accordance with the definition of a viewpoint, any reasonable set of architectural concepts and structuring rules may be chosen to focus our attention on particular concerns within a system; and therefore we often use a *business viewpoint* and an *information system viewpoint*[4]. In this manner, for example, we can exactify the familiar and useful slogan "no requirements in terms of solutions" since requirements and solutions belong to different viewpoints.

Only primitives? RM-ODP provides a small but powerful system of interrelated definitional primitives that you can use to build your own specification. These primitives drastically reduce the number of base things and relationships and, hence, the complexity and size of a specification. More importantly, they *reduce the number of concepts* that the readers of a specification have to master in order to understand it.

The primitives need to be expanded in actual specifications. RM-ODP helps here in the form of "structuring rules" specifically designed to allow RM-ODP primitives to be used to develop more complex and/or specific definitions of various *business patterns*. These specialized definitions can be successfully mixed with the original primitives to create increasingly rich systems of definitions. This is similar to the way in which mathematicians create arbitrarily rich theorems from other theorems and well-understood basic axioms, or similar to how engineers create arbitrarily large and complex structures from common subassemblies. In creating new definitions we should not forget that:

- a warm and fuzzy feeling is not a definition
- a "meaningful name" is not a definition by itself
- an example is not a definition, and
- requirements are not the same as solutions (and should not be formulated in terms of solutions).

The business patterns, of course, have to be discovered and explicitly formulated — and this is what business domain modeling is for. It discovers and specifies *deep analogies between seemingly different things, relationships, and processes*. In this manner, organizations can understand and deal with "always-changing" requirements as variations of a small number of conceptually simple patterns, leading to substantial savings in intellectual effort, time, and money. Among other things, it becomes possible to be demonstrably *proactive* rather than reactive in solving business problems because a clear and crisp business model may by itself provide a substantial compet-

---

[4]This distinction between business and system viewpoints was made explicit, among others, in the distinction between computation-independent and the other two (platform-independent and platform-specific) viewpoints of the OMG's Model-Driven Architecture.

itive advantage for the modeled enterprise. To quote a satisfied client from a large financial firm, "[i]t has changed the way the client views software development and this single effort will serve as the foundation for other planned software development initiatives. This business specification, written for software development, has potential application in other areas. Portions of the specification can be incorporated in corporate policy manuals; regulatory compliance documents, and serves as a basis for business process review." [G2001] In other words, various business and IT decisions could be based on a solid and explicit foundation rather than on handwaving, eloquence of gurus, or lemming-like considerations. Similarly, to quote a satisfied customer from the pharmaceutical industry, this approach is "thorough and flexible enough to specify the work-products and linkages [relationships] of architecture at any level of abstraction" as opposed to "some fuzzy, ill-defined and confused notions".

Ontologies and invariants

Domain semantics. When we want to use an existing system (component) or plan to use a new one, we need to specify (for existing systems it often means "reverse engineer") the semantics of its interfaces, since only precise and explicit semantics make interoperability possible. This applies to any kind of system, independently of whether it is (or will be) computer-based. Clearly, in order to understand and define (the requirements for) interface semantics, we need to be *explicit* and precise about the business domain in which the system works or will work. This is true for all kinds of businesses — be they traditional ones such as trading, or the business of creating an information management system, or of creating and using a relational database, or of asynchronous messaging, or of a particular general ledger legacy program. In other words, understanding of "what is there in the business domain" (also known as ontology) comes first. Within this explicitly specified framework, we can discuss the systems (and their interfaces) that work or are planned to work in this domain (after all, the descriptions of these systems refer to the things, relationships and actions of the domain!). And a data dictionary is *not* such a framework: as noted by many and as we know from practice, the same name may mean substantially or somewhat different things in different contexts, and the *structure* of these contexts — expressed in the relationships — is essential for understanding of the things we deal with.

It is now well understood that attempts to comply with a specification having incomplete or unclear semantics will not guarantee interoperability because important information will be lost to handwaving. As noted earlier, RM-ODP and GRM provide concepts that make it possible to completely and precisely — i.e., formally — specify a system within its environment, be it a business or an information management system. These definitions are based on the semantics of the appropriate *domain* rather than on existing systems, products, or solutions. Similarly, specifications of existing products, including legacy systems, also can and should be provided in terms of concepts and constructs from RM-ODP.

Business (and IT system) domain modeling is a branch of software engineering. Of all disciplines, software engineering is unique in handling all aspects of extremely complex artefacts (created by humans) and thus in practical, engineering, usage of abstraction and exactification. Unlike other disciplines dealing with complex artefacts, software engineering has to model and reason about,

using the same basic concepts and constructs, all abstraction levels of an artefact such as a specification or a program — from top-level strategy to bits. This uniqueness was noted by Dijkstra, Bjørner, and other classics of computing science.

It is very instructive to observe that the underlying concepts for ontology understanding, development and use in business and system analysis are not radically new. They have been known from exact philosophy (notably, Bunge's work [B2003, B2004]), database and object modeling work based on Bunge (see, for example, [WW2004]), systems thinking (F.A.Hayek[5], von Mises and others), and, of course, such international standards as RM-ODP and GRM. These are the same few basic concepts, and more often than not they have been defined in the same manner. The concept of composition is among the most important, and its definition — based on levels of abstraction and on emergent property determination — is essentially the same in RM-ODP, GRM, and in Bunge's work. Much more details and a large number of examples are provided in [K2002].

The importance of elegance. With the introduction of standards like RM-ODP and GRM, today the analysis situation resembles the one that existed in programming in the second half of the 1970s. At that time, in E.W. Dijkstra's words, programming was in the process of moving from a craft to a scientific discipline. Most observations made by Dijkstra then for programming are often reinvented now for analysis. We also see that in analysis, as in programming (an observation made by Dijkstra as early as 1962), elegance is of utmost importance; elegant specifications are liked by all and, thus, successfully used and reused. As Dijkstra said, "in the design of sophisticated digital systems, elegance is not a dispensable luxury but a matter of life and death, being a major factor that decides between success and failure", and a good specification ought to be convincing in the same manner as a good program is. For a specification to be of use (and to be produced in a reasonable manner, as any artefact produced by professional engineers), it needs to describe semantics (meaning) rather than syntax, and to do that in an abstract and precise manner. This is precisely what RM-ODP and GRM have been designed to do.

"Learn to abstract: try not to think like a programmer" (J.Wing). Unfortunately, both in programming and in analysis, the fashionable approaches had often encouraged practitioners to start their work in the middle using an operational approach. Jeannette Wing provided an excellent example of the approach to be avoided: "'If you do this and then that and then this and then that, you end up in a good state...' This [...] process quickly gets out of control. The problem is related to understanding invariants." [W1996]. *Invariants* are essential for defining the ontology of the domain of interest, be it a business or an information system one. In particular, they define the types of things, actions and relationships. They also determine what actions (and under what circumstances) can and cannot be executed in the context of the domain ontology. To quote a recent eloquent paper by Turski, one of the founders of computing science, a fundamental breakthrough in programming happened when "[i]nstead of preoccupation with a dynamic process ('what happens next'), we concentrated on a piece of text ('what does it say')" [T2003]. Indeed, the same kind of difference exists between a (failed) procedural approach to teaching and

---

[5]For example, in accordance with Hayek's observations, prices constitute an essential emergent property that makes possible the functioning of a market economy and that embodies more information than each participant of a market economy directly has.

learning mathematics (or computing science) and an approach based on understanding. And the same kind of difference exists between various buzzword-compliant operational approaches to analysis that quickly get out of control, and elegant approaches that lead to understanding of businesses and information systems.

These elegant approaches start with and are based on ontologies. We do not want to start in the middle (that is, with a possible solution, as it too often happens) or even with a specific problem (that is, with requirements). Rather, *we start with the stable (that is, invariant) basics* and proceed from there. Clearly, a problem and its solution cannot be understood and specified without the basics because the interrelated concepts used in describing the problem and the solution are defined by (and in) the basics.

Discover different ontologies. Substantial and especially somewhat different ontologies of different stakeholders (especially tacit ontologies) may make semantic interoperability difficult to achieve. Each business department and system has its own special, usually implicit, vocabulary. Working together in an electronically connected world means effectively integrating people, businesses, and their systems, and this depends on being able to elicit their special vocabularies and translate between them. In order to solve this problem, the existence of different ontologies has to be made explicit. Ontologies — business domain models — are the framework for inter-relating these existing vocabularies, not throw them away, and provide interoperability solutions that really work, because they involve business *meanings*, while purely technical (or syntactical) solutions fail. Explicit ontologies make it possible to discover and specify mappings between concepts and relationships used in different ontologies (and often to enrich some of them). As a result, different stakeholders can communicate and interoperate because it will become clear which different context-specific terms have the same meaning and which identical context-specific terms have different meanings.

"Requirements always change"? In this context, it is instructive to elucidate the (all too familiar) statement "requirements always change and therefore it is useless to formulate them". Indeed, business *processes* often change. Such changes may lead to a competitive advantage for a business or they may even be perceived as necessary for the business to survive. Similarly, decisions about using IT systems to automate certain business processes may also change (for example, due to perceived opportunities). At the same time, the basics of a business — its ontology — have usually remained the same for centuries, if not for millennia: for example, banking and financial texts published in the early 20th century (or earlier, such as fragments from Adam Smith's *Wealth of Nations*) have been successfully used to understand and specify the corresponding business *domains*. The changes due to modernity are minimal and are mostly additions or refinements to the existing classical models.

These considerations apply to any kind of business modeling as well as to requirements discovery and specification, independently of whether a computer-based IT system will be created or bought to automate some business processes or steps. As noted earlier, a crisp business model is used *to make demonstrably effective business decisions* only some of which are IT-related. And the concepts, constructs and standards used in creating such models are based on mathematics, "the art and science of effective reasoning" (Dijkstra).

Notations

Both in traditional programming and in modeling, we know from the works of Dijkstra and other founders — as well as from the experience of the best practitioners — that the inherent complexities of the problems and their solutions should not be exaggerated by imposing on the readers of programs or specifications a complex notation "with a plethora of ad hoc facilities of dubious value and unquestionable ugliness" [T2003]. At least in traditional programming, program readers and writers are usually humans with roughly the same background in the notation used. Contrariwise, in modeling (and specifications — the result of modeling), readers and writers often have drastically different backgrounds in the notations used. Therefore in order for business experts to use a precise notation to communicate about business models, it is absolutely essential to explain the basics of the notation on the 'back of an envelope'. This is important to get, and retain the attention of the business people. Imposing a very complex notation will not make the problems simpler, and will, in fact, move the reader's (and modeler's) attention from the complexity of the problems to the complexity of their representation. In other words, the attention will be moved from the essential semantics to the accidental syntax. As a result of such semiotic pollution, communication about problems suffers. This explains why many business system specifications are write-only, at least from the viewpoint of their main customers — the business decision makers.

Business (and any system) modeling notations should not be restricted by the artefacts existing or easily implementable in currently used IT systems,. As an example, such a notation ought to be able to express multiple and dynamic subtyping, multiple decompositions of the same individual, non-binary relationships with well-defined semantics, and so on. (If these concepts are realized by IT systems (existing or to-be-built) and if the IT system is chosen to be object-oriented then there ought to be a mapping from a high-level implementation-independent construct to a set of lower-level, technology-specific constructs.) For example, it should be possible to specify that a banking industry is a composite in a composition[6] of banks, a federal regulator, customers, and a lot of something else (like clearing houses). Moreover, it should be possible to say that in this composition both the composite and the components ought to exist together. The specific number of components may also be determined: say, exactly one regulator, at least two banks, and so on. And of course we need to be able to specify that this composition exists!

Turski stresses the need to limit ourselves in the process of understandable program construction

---

[6]A composition (also sometimes known as a UML aggregation — see more details about this aspect of UML in [EDOC2001]) is a relationship between a "whole" (composite) and its "parts" (components). The type of the "whole" corresponds to the types of one or more "parts", and an instance of the "whole" corresponds to zero or more instances of each type of the "part". There are two kinds of the properties of the "whole": those that are determined by the properties of the "parts" and the way these "parts" are combined; and those that are independent of the properties of any "parts". A composition also satisfies the general relationship invariant that implies, in particular, that an instance of the "whole" cannot have itself as a "part". This definition is based on the definition of composition in RM-ODP, GRM, and on the definition of composition used in systems thinking (for example, by F.A.Hayek) and in exact philosophy (or example, by Mario Bunge).

to the systematic use — known as "structured programming" (invented by Dijkstra) — of a small collection of programming language instructions having "a clean and well-defined meaning". Further, he emphasizes that a high quality specification has to have a model not only in the programming language domain, but also in the language "used for the description of (a part) of the reality of interest". In other words, this later language should be understandable to the business subject matter experts. Following the ideas of structured programming, we may want to limit ourselves in the process of understandable specification construction to the systematic use of a *small collection of modeling constructs having a well-defined meaning*. And this collection, in fact, this system, already exists and was described in RM-ODP.

When we have to choose or use a specific notation or tool in our programming or modeling activities, we should, first and foremost, look at whether and how the semantics of concepts we use in programming or modeling is supported by the tool. If the notation or tool is overwhelming then everything need not be lost: it may often be possible to choose a (very) small subset of that notation in order to represent concept semantics. This approach is not new at all: it is well-known, for example, that various small subsets of PL/I have been created and used exactly for this purpose. Similarly, a very small subset of UML for business modeling has been created and used in various engagements such as that described in [G2001] and in several papers in [KB2003], as well as for specifying relationships in [EDOC2001]. This subset has been represented on one page.

At the same time, an important caveat is in order. When a subset of a notation is being chosen to represent concept semantics, it is essential for the semantics to have an exact representation in the notation. Since many important aspects of UML semantics have not been well-defined, it became necessary to provide for precise definitions of UML constructs used to represent the semantics of such concepts as composition, subtyping and reference relationships. This approach was accepted by OMG in the UML profile for EDOC [EDOC 2001].

Business patterns and modeling

The reference list includes books and papers with fragments of business and IT system specifications based on RM-ODP and GRM. None of these are toy examples. Some examples from textbooks are illustrative and fun but not trivial — like those modeling fragments of domains described by Lewis Carroll — while others are fragments from the generic parts of industrial specifications created for (and with active participation of) business and IT customers. These generic business and IT specifications may be, and have been, reused as business patterns in various customer engagements: after all, *pattern matching in context* is an essential part of successful analysis (and design). Since generic business patterns — such as, at different levels of genericity, invariant, composition, or contract— can be used in any application area, a good analyst can become a contributor in an entirely new area within a very short timeframe. The conceptual foundation together with the generic business patterns let the good analyst to ask proper questions and "begin speaking the language [of the entirely new area] competently within a week or so" [W1982]. Clearly, this systems thinking (and thinking in general) approach differs from that of the authors of many "help wanted" advertisements.

And where do the business-specific business patterns come from?

Most of the concepts of a business, whether financial services, document management, or telemedicine, are common to the entire community of that business. We capture these domain concepts most effectively by reusing and improving their definitions over the decades (especially since some of them have been around for centuries, see, for example, Adam Smith's *Wealth of Nations*), not by rewriting and redefining them as "requirements" for each engineering project, taking precious time from business people and programmers to do so, and in too much of rush to get it right. In other words, ontology reuse — and concept reuse in general — is much more valuable than code reuse.

Let us recall that Peter Naur proposed in 1968 [SE1969] to use the work of Christopher Alexander long before it became fashionable to refer to it as a source of ideas about attacking the software design problem. Naur justified his choice by the fact that Alexander was concerned with the design of large heterogeneous constructions. Indeed, Alexander emphasized in *The Timeless Way of Building* that "...a pattern defines an invariant field which captures all the possible solutions to the problem given, in the stated range of contexts... the task of finding, or discovering, such an invariant field is immensely hard... anyone who takes the trouble to consider it carefully can understand it... these statements can be challenged because they are precise" [A1979].

Creating and elucidating a domain model cannot be automated. There is no algorithm (or tool) to do that. Such models are created and elucidated by teams consisting of domain SMEs (subject matter experts) and analysts. Even when the SMEs are experienced in formulating ontologies of their domain in an abstract (that is, understandable) and precise manner, analysts are still essential to discover, elucidate and exactify tacit assumptions common to all, or some, SMEs. *Ignorance* — real or perceived — of the subject matter is needed to make the tacit assumptions explicit. As early as 1969, P. Burkinshaw urged: "Get some intelligent ignoramus to read through your documentation; [...] he will find many 'holes' where essential information has been omitted. Unfortunately intelligent people don't stay ignorant too long, so ignorance becomes a rather precious resource." [SE1970].

It is not sufficient to discover, formulate and use (fundamental, basic, and more specific) concepts and structures essential for a good model. We ought to communicate these discoveries, both for understanding of that model and for their usage in other, often apparently very different, models. In the modeling context, a concise and elegant system of basic concepts described in RM-ODP provides a foundation for such a language. Sometimes the *specifics* of this language or, more often, its fragments ought to be created (collectively, by the modelers together with the subject matter experts) for successful communication of a model's semantics. Models formulated in such a manner (of course, based on concepts and structures common to most systems) establish a common background used by all stakeholders of an organization and its relevant environment (e.g., clients and subcontractors) in understanding and business decision making.

Conclusions

It is now well understood that attempts to comply with a specification of any system having incomplete or unclear semantics will not guarantee interoperability because important information will be lost to handwaving. The system of concepts defined in RM-ODP and GRM makes it possible to completely and precisely — i.e., formally — define the universe of discourse, be it to describe a business or an information management system. These specifications are based on the semantics of the appropriate domain rather than on existing systems, products, or solutions. Since the system of concepts and constructs used for these specifications has been itself defined in a clear and crisp manner, the specifications can be *read*, understood, and thus agreed or disagreed upon by all stakeholders. Moreover, specifications of existing products or systems — including legacy systems — can and should be provided using the same approach and thus leading to demonstrably justified user decisions about acquiring such systems.

An RM-ODP-based specification provides a top-level precise (not semi-precise!) road-map of the appropriate fragment of a business or IT system, or of a product, and with its refinements down to the level(s) we are interested in. These specifications define what should always be true about the things and relationships of the business or IT domain as well as what should be true about each process (step, operation). All defaults are made explicit: in particular, the developers do not need to ask "this person in the corner" since everything they need to know about the business domain and "were afraid to ask" is in the specification. Such specifications may be, and have been, used for making demonstrably justified strategic, tactical, and operational decisions in all kinds of business and IT system environments.

How to proceed? Look at the reference list. You may want to start with the Cutter Consortium publications about RM-ODP [KG1999] and business models [K2001] and continue with the OMG's Semantics Working Group Green Paper [KT1997] and the business modeling text [K2002]. The articles quoted above, in the subsection "Is RM-ODP really useful?", provide some specifics about the usage of RM-ODP and GRM in various industrial environments. Other references include interesting, useful and often fun background material.

The "back to basics" approach helps a lot!

Acknowledgments

Many thanks go to the colleagues and customers the interactions with whom helped to distill the concepts and approaches described here. Some of their names are in the reference list below, and many names are in books and papers from this list. Also, specific thanks go to Rashid Bashshur for useful discussions about the structure of this paper.

References

[A1979] Christopher Alexander. *The Timeless Way of Building*. Oxford University Press, 1979.

[B1873] W. Bagehot. *Lombard Street: A Description of the Money Market.* Scribner, Armstrong & Co., New York, 1873.

[B2003] Mario Bunge. *Philosophical dictionary.* Prometheus Books, 2003.

[B2004] Mario Bunge. *Emergence and convergence.* Toronto University Press, 2004.

[D1986] E.W. Dijkstra. *Management and Mathematics.* EWD966, The University of Texas at Austin, 14 June 1986.

[EDOC2001] A UML Profile for Enterprise Distributed Object Computing Joint Final Submission Part I. 18 June 2001. OMG Document Number: ad/2001-06-09. 3.6. The Relationship Profile.

[G2001] James S. Garrison. Business specifications: Using UML to specify the trading of foreign exchange options. *Proceedings of the 10th OOPSLA workshop on behavioral semantics (Back to Basics)* (Eds. K. Baclawski and H. Kilov). Northeastern University, Boston, 2001, pp. 79-84.

[GRM] ISO/IEC JTC1/SC21, Information Technology. Open Systems Interconnection - Management Information Services - Structure of Management Information - Part 7: General Relationship Model, 1995. ISO/IEC 10165- 7.2.

[H1957] Grace Hopper. *Automatic Programming for Business Applications.* In Proceedings of the 4th annual computer applications symposium, October 24-25, **1957**, Armour Research Foundation, Chicago.

[HE1999] J.Hassall, J.Eaton. Applying ISO RM-ODP in the specification of CORBA interfaces and semantics to general ledger systems. *Behavioral specifications of businesses and systems* (Ed. by H.Kilov, B.Rumpe, I.Simmonds), Kluwer Academic Publishers, 1999, pp. 91-104.

[K1999] H. Kilov. *Business Specifications.* Prentice-Hall, 1999.

[K2001] H.Kilov. *Anticipating the market: The value of business models.* Distributed Enterprise Architecture Advisory Service Executive Report. Cutter Consortium, 2001.

[K2001a] Thomas Kudrass. Coping with semantics in XML document management. In *Proceedings of the 10th OOPSLA workshop on Behavioral Semantics — Back to basics.* (Tampa, Florida, 15 October 2001), pp. 150-161.

[K2002] Haim Kilov. *Business models.* Prentice-Hall, 2002.

[KA1997] H.Kilov, A.Ash. How to ask questions: Handling complexity in a business specification. *Proceedings of the OOPSLA'97 Workshop on object-oriented behavioral semantics* (Atlanta, October 6th, 1997), ed. by H.Kilov, B.Rumpe, I.Simmonds, Munich University of Technology, TUM-I9737, pp. 99-114.

[KB2003] H.Kilov and K.Baclawski (Eds.) *Practical foundations of business system specifications.* Kluwer Academic Publishers, 2003.

[KG1999] H.Kilov, M.Guttman. *ISO Reference Model for Open Distributed Processing: an informal introduction.* Cutter Consortium (Distributed Computing Architecture Advisory Service) Executive Report, Vol. 2, No. 4 (April 1999). ISSN 1523-5912..

[KMS1996] H.Kilov, H. Mogill, I. Simmonds. Invariants in the Trenches. *Object-Oriented Behavioral Specifications* (Ed. by H. Kilov and W. Harvey). Kluwer Academic Publishers, 1996, pp. 77-100.

[KR1994] H.Kilov, J.Ross. *Information modeling.* Prentice-Hall, 1994.

[KT1997] H.Kilov, K.P.Tyson. *Semantics Working Group Green Paper One.* OMG Semantics Working Group. OMG Document Number ormsc/97-06-10r.

[L1990] Yuri M. Lotman. *Universe of the mind: a semiotic theory of culture.* Translated by Ann

Shukman. Introduction by Umberto Eco. London and New York: I.B.Tauris & Co. 1990.

[L2001] Herbert W. Lovelace. The medium is more than the message. *Information Week*, July 15, 2001.

[LQS] http://informatics.mayo.edu/index.php?page=11

[R1999] K.Riemer. An analysis of RM-ODP viewpoints and system life cycles. In: *Proceedings of the 8th OOPSLA workshop on behavioral semantics* (Ed. by K.Baclawski, H.Kilov, A.Thalassinidis, K.Tyson). Northeastern University, 1999.

[RM-ODP] ISO/IEC JTC1/SC21. Open Distributed Processing - Reference Model: Part 2: Foundations (ITU-T Recommendation X.902 | ISO/IEC 10746-2). This part is essential.

[SE1969] *Software Engineering*. Report on a Conference sponsored by the NATO Science Committee, Garmisch, Germany, 7th to 11th October 1968. (Chairman: Professor Dr. F.L. Bauer, Co-chairmen: Professor L. Bolliet, Dr. H.J. Helms; Editors: Peter Naur and Brian Randell). January 1969.

[SE1970] *Software Engineering Techniques*. Report on a Conference sponsored by the NATO Science Committee, Rome, Italy, 27th to 31st October 1969. (Chairman: Professor P. Ercoli, Co-Chairman: Professor Dr. F.L. Bauer, Editors: J.N. Buxton and B. Randell) April 1970.

[SKB2001] Lawrence E. Sweeney, Enrique V. Kortright and Robert J. Buckley. Developing an RM-ODP-based architecture for the Defense Integrated Military Human Resource System. *Proceedings of the WOODPECKER-2001 (Open Distributed Processing: Enterprise, Computation, Knowledge, Engineering and Realisation) in conjunction with ICEIS'2001*, Setubal, Portugal, July 2001, pp. 110-123.

[SW2004] Hewlett-Packard. Introduction to Semantic Web technologies. http://www.hpl.hp.com/semweb/sw-technology.htm

[T2003] W.Turski. It was fun. *Information Processing Letters*, 88(2003), 7-12.

[TS1998] A.Thalassinidis and I.Sack. Building the Industry Library — Pharmaceutical. In: *Second ECOOP Workshop on Precise Behavioral Semantics (with an Emphasis on OO Business Specifications).* (Eds. H. Kilov and B. Rumpe), Technische Universität München, TUM-I19813, June 1998, pp. 245-253.

[W1982] Gerald Weinberg. *Rethinking systems analysis and design.* Little, Brown, and Company, 1982.

[W1996] Jeannette M. Wing. Hints to specifiers. In: *Teaching and learning formal methods.* (Eds. C. Neville Dean and Michael G. Hinchey). Academic Press, 1996, pp.57-77.

[WW2004] Yair Wand, Ron Weber. Reflection: Ontology in information systems. *Journal of Database Management*, 15, 2 (April-June 2004), iii-vi.

Semantic Interoperability in Telemedicine through Ontology-driven Services

Pronab Ganguly* & Pradeep Ray+
*School of Computer Science and Engineering,
+School of Information Systems & Management
The University of New South Wales, Kensington,
NSW 2052, Australia,
*Tel: 61 2 96915762, Email: pganguly@qantas.com.au
+Tel: 61 2 9385 5890, Email: p.ray@unsw.edu.au

1.0 Introduction

Telemedicine involves the integration of information, human machine and healthcare technologies. As different modalities of patient care require applications running on heterogenous computing environment, software interoperability is a major issue in telemedicine. Software interoperability may be defined as"the capability with which two or more programs can share and process information irrespective of their implementation language and platform" [1]. Our study [2] outlined the development methodology for software agent based interoperable telemedicine systems. Subsequently Ontology has emerged as a major technique for software interoperation. In this paper the importance of ontology will be highlighted from the perspective of semantic interoperation. Section 2.0 describes the issues in software interoperation. Section 3.0 highlights the importance of ontology in software interoperation. Section 4.0, 4.0, 6.0, 7.0 & 8.0 describes the major element of our framework. Section 9.0 describes some related efforts and Section 10.0 describes the scope of further research in this area.

2.0 Software Interoperability

Software interoperability may be defined as the ability for multiple software components to interact regardless of their implementation programming language or hardware platform. The available mechanisms for software interoperability are:

- Physical interoperability: In this approach, the interoperability is achieved by physically transferring the information through electronic media such as floppy disks, CDs and magnetic tapes.
- Data-type interoperability: distributed and disparate programs support structured exchange of information through Application Programming Interfaces (APIs) invoked over a computer network.
- Specification-level interoperability: same as the previous one but also encapsulates knowledge representation differences at the level of abstract data types (e.g. a Table, Tree etc.). This enables programs to communicate at higher levels of abstraction and increases the degree of information hiding. CORBA and Enterprise Java Beans (EJB) fall into this category.

- Semantic interoperability: unlike the above two types of interoperability which are concerned with the form (structured description) at the integration interface, semantic interoperability represents design intent and predicted behaviour as well as form (structured description) of the shared entities. It assumes that different information sources store information on related issues but each may offer a different meaning (semantics) of it.

The semantic interoperability problem may be defined in general as "the ability of a user to access, consistently and coherently, similar (though autonomously defined and managed) classes of digital objects and services distributed across heterogeneous repositories, with federating or mediating software compensating for site-by-site variations"[3].

Thus semantic interoperability between various heterogeneous information sources continues to pose serious challenges to database, artificial intelligence and other related communities. This issue includes system, syntactic and structural/schematic interoperability. When we restrict the context to a scenario where various domain-specific but diverse databases are simultaneously accessed, semantic interoperability may be achieved. However, even in such cases, these databases are heterogeneous in the sense that they store different types of data, different data representation formats and there are different software and computing platforms used to run these databases. Researchers are working towards the objective of identifying the mechanisms that can access data from multiple databases without making changes to the existing databases. Existing approaches for integration of heterogeneous databases include resolution of structural differences between underlying databases. In conceptual or global schema approach, there is a need for standardisation of data structures and definitions [4]. The conceptual schema specifies field and record definitions, structure and rules for updating data values. Various mappings and transformations are used to convert source data into a semantically equivalent and compatible form. The problems with this approach include:

- The emphasis is on schematic (i.e. syntactic) rather than semantic heterogeneity
- It assumes global knowledge is available which is not always possible
- The development costs of coding the system and the data definitions is huge
- The autonomy of individual databases is lost

In a federated database approach, a multidatabase language is used to facilitate interoperability. This language is the vehicle for interoperability. Additionally, each local database provides an export schema (a portion of its overall schema) which it is willing to share with other sources. Each database then uses these export schemas to define an import schema. Thus a partial global schema representing information from other remote databases is generated and used. But the problem is that the users need to determine the meaning of all existing concepts and terms in every database across the network. This limits the number of databases and thus the scalability even in the same domain. It is very difficult to represent semantic information outside a specific domain. Even in a specific domain, human (user) intelligence is usually required to be aware of the context so as to assess semantic information such as in a federated approach. The issues include:

- A specific action can have different results depending on the context. For example, in a telemedicine application, the action "follow a normal diet" will mean different things to different users.

- When using a global schema, it is hard to maintain the autonomy of individual databases and keep the overall system's development costs as low as possible.
- Application-specific solutions that compromise the autonomy, flexibility and scalability of the entire distributed application should be avoided.

Our solution is based on the concepts of *Ontologies* and software agents. The major elements of the proposed system are:

- Since contexts need to be used to represent the underlying semantics of the databases, these contexts are defined in ontology. . So the users only need to express their needs by using terms in the ontology.
- There are *software agents* collaborating with every (human) user in the distributed environment as well as the ontology and the databases. These software agents are considered like human entities (having beliefs, desire and commitments) constrained by pragmatics (intentions, communication etc) as well as semantics (meanings, propositions, validity etc) and syntax (formal structure, data etc). They handle all communication aspects, receive feedback from the user, initiate communication, monitor events and perform certain tasks.

3.0 Ontology and Semantic Interoperation

The word ontology is derived from the Greek words "ontos" (which means "being") and "logos" (which means "word"). Thus ontology is related to various kinds of things that may exist in a given domain. Basically ontology provides a formal specification of the terms in a given domain and relationship among them. Thus Ontology is a formal, explicit specification of a shared conceptualization [5].

- "A 'conceptualization' refers to an abstract model of some scenarios in real life which identifies the relevant concepts of that phenomenon."
- "'Explicit' means that the type of concepts used and the constraints on their use are explicitly defined."
- "'Formal' refers to the fact that the ontology should be computer readable."
- "'Shared' reflects the notion that an ontology captures consensual knowledge, that is, it is not restricted to some individual, but accepted by a group".

A knowledge base consists of an ontology and instances of the associated classes. As the built-in knowledge in ontology is consensual in nature, it can be shared. In the practical sense, development of a domain ontology involves:

- Identification and definition of classes
- Arrangement of the classes in a taxonomical hierarchy
- Attaching properties (slots) and defining allowed values (facets) for these slots
- Allocation of the values for slots for instances

As ontologies are described in a logic based language, the usual features of logic are available. So accurate, distinct and logical linkage as well as distinction can be made among the classes, properties and relations. This helps inference engines o conduct automated reasoning with the help of ontologies. Similarly support is provided to intelligent entities such as software agents. Ontology is a major element for representation of the semantics of the document. Thus ontology can be effectively

used as a mechanism for semantic interoperation in telemedicine as well. In order to illustrate this aspect of ontology, we will first describe a telemedicine application in diabetes management and then highlight importance of ontology as a means for semantic interoperation.

4.0 Telemedicine and Diabetes Management

Diabetes is a condition in which the level of blood glucose is persistently raised above the normal range. If the condition is left untreated it can lead to renal and ocular complications or damage to peripheral nerves [6]. Worldwide, 135 million people suffer from diabetes. The figure is projected to grow by 122% within the next 25 years. These statistics clearly indicate the growing need for an efficient and effective treatment plan for those who suffer from diabetes.

Currently, treatment of diabetes entails limited patient contact with multiple healthcare professionals such as the general practitioner (GP), specialist and clinicians. The interaction would involve regular patient visits to the family GP who collect blood samples to send to the clinical laboratory for testing. The result is reviewed by the GP to determine the patient's condition to prescribe medication. All complicated cases are referred to a specialist.

The above interaction outline the type of organised care currently provided to the diabetic patient. Two major problems identified in the current system are as follows:

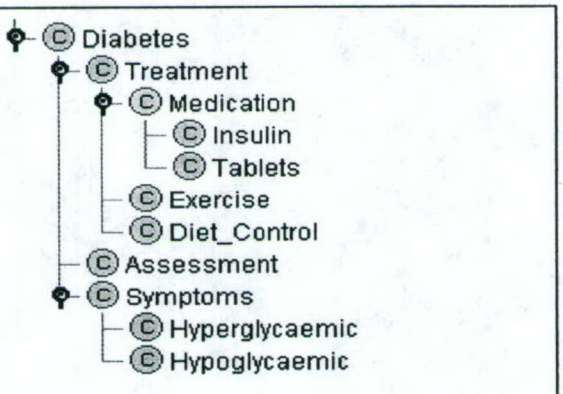


Fig. 1. Simple conceptual definition of the diabetes domain.

- The patient, laboratory clinicians, GP and the specialist interact effectively amongst each other to provide efficient health care to the patient.

- A new approach will be established to maintain consistency within a heterogeneous environment.

We are proposing an ontology-driven multi-agent system for diabetic treatment. This system aims to manage complex human interactions by the use of a diabetes ontology and agent system.
The major components of this implementation are:

- *FIPA Standard based Ontology* development using Protege.
- Toshiba Beegent Platform

The next section presents the need for an ontology-driven agent framework for diabetic management.

5.0 Ontology Driven Software Based Framework

A software agent is a program which performs a specific task on behalf of a user, independently or with little guidance. A software agent also facilitates the communication of programs written in different languages. In the diabetes scenario, each healthcare professional may use different systems to assist in management of patient records and treatment plans. Each system may reside on different platforms across a network. To unify these systems effectively, a software agent could be used. The agent would create a "wrapper" around each application and communicate with other wrappers using an agent communication language. Therefore the healthcare professionals will benefit from an agent-based system to support existing healthcare frameworks.

The behavioural constraints on software agents can be defined so that the agent can change behaviour based on other dynamic components within the system. For example, agent technology can be used to enhance doctor-specialist collaboration. A typical doctor-specialist interaction would involve:

- Doctor contacting specialist about patient
- Specialist formulating treatment plan based on patient information
- Doctor formulating treatment plan based on patient information
- Doctor and Specialist negotiating ideal treatment plan for patient

An agent system can be used to try and emulate part of this behaviour independently of the real doctor or specialist. To effectively emulate this behaviour requires not only the development of an agent, but also the development of ontology. Ontology is a formal, explicit specification of a shared conceptualisation as defined in. It provides a vocabulary of terms and relations with which to model the domain. Ontology is suited to represent high-level information requirements to specify the context information in a collaborative environment.

The ontology will help to share common understanding of the structure of information. Agents can then be used to extract and aggregate information to answer queries to other applications. In the above interaction, if the doctor uses the term "drug" to indicate the medicine that a patient is taking, and the specialist uses the term

"treatment" to indicate the same term, a simple multi-agent framework will not be able to interpret the context of each message although it has the same meaning. To facilitate the conceptual understanding of terms, ontology should be developed. An agent can then query this ontology to perform an intermediate translation of terms for use by other agent systems.

Fig 1 describes a simple conceptual definition for the diabetes domain. Classes (also called concepts) are the main focus of ontologies. Classes are used to describe concepts in the domain. A class can have subclasses that represent concepts more specific than the superclass. In Fig 1 Diabetes represents the main concept containing specific sub concepts such as Treatment and Symptoms. The Treatment concept can be further refined to represent Medication, Exercise, and Diet Control. By defining relationships and constraints for these concepts, we would be able to formally describe the diabetes domain. Once fully developed into a knowledge base, an agent can access these terms and relationships to reason and to derive answers to queries. An ontology driven multi-agent system can then be used to resolve the issue of maintaining consistency with medical terminology and standards.

6.0 Architecture of an Ontology-Driven Agent System

The conceptual model of an ontology-driven agent system is shown in Fig. 2.

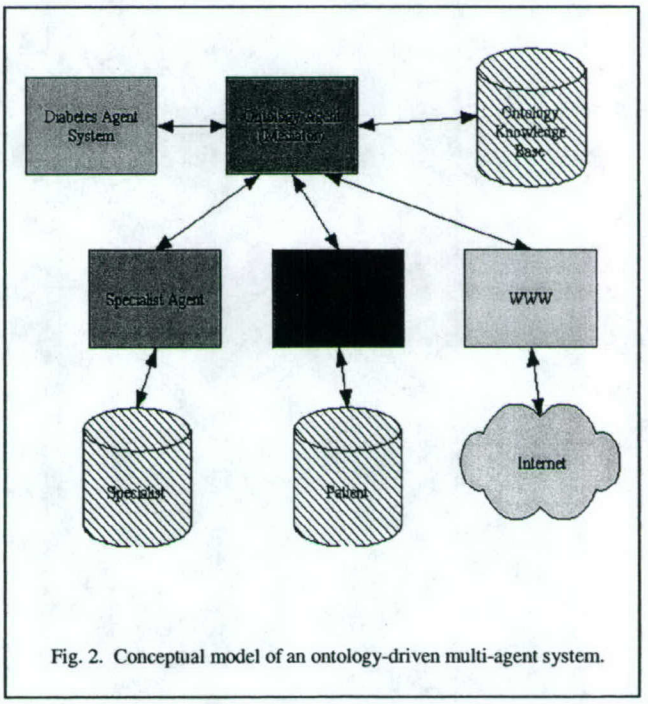


Fig. 2. Conceptual model of an ontology-driven multi-agent system.

The model is based upon an:

- Establishment of agent systems
- Interaction via an agent communication protocol

- Development of an ontology

This model will be in compliance with the Foundation for Intelligent Physical Agents (FIPA) specification explained in the next section.
The following agent systems will be established:

- Diabetes Agent System
- Ontology Agent
- Specialist Agent
- Patient Agent
- WWW Agent

Each agent performs a specific set of function as stated below:

Diabetes Agent System:

- Provides an interface for the GP or specialist. Should incorporate an Eliza based algorithm to respond to the user in natural language.
- Sends any requests from the user to the Ontology Agent (OA).
- Receives messages from the OA to display, or to respond to.

Ontology Agent:

- Receives messages from any agent on the system.
- Messages are broken into tokens to be translated into primary concepts. This interaction takes place in collaboration with the Knowledge Base. For example if a message contains a token called "web" and the primary concept defined in the ontology is "Internet", all messages containing the token web will translated into the token Internet.
- OA then queries all agent systems to determine which agent can satisfy the translated message. The OA then sends the message to the destination agent.

Specialist Agent:

- Provides information about treatment options for a patient. Has access to a knowledge base to determine correct treatment options based on symptoms.
- Can negotiate with Diabetes Agent System to arrive at optimal treatment plan.

Patient Agent:

- Responds to request for patient information. Example, blood glucose levels, symptoms, general patient data
- Should interface with a patient to gather information. Could also incorporate an Eliza based algorithm

WWW Agent:

Provides access to the World Wide Web. Performs a search on data requested by other agents and launches an interface with the requested information.

A scenario of how this system will function is as follows:

The user (GP or specialist) would request for information to the Diabetes Agent System. This request would be sent to the Ontology Agent to be translated based on the Ontology developed. An example request could be to find information from the Internet about diabetes. The ontology agent would traverse each word in the message and translate it into a key concept. For example, the term "web" can be translated into the primary term Internet. The new message would then be: "find information from the Internet about diabetes". This is done so that all messages passed within this system are consistent and can be understood by all agents.

After message translation, the ontology agent has to determine which agent can understand and perform the request. To do so, the ontology agent will query each agent to determine the one that satisfies the key terms in the message. If there is a match then the message is passed to the agent.

For example, the message: "find information on the web about diabetes", would be sent to the WWW Agent. This search agent will then launch a browser to find information on diabetes.

The advantages of developing a system this way is that we are not hard-coding "trigger words" to activate certain actions. Rather we are developing agents to deal with certain concepts within a fixed domain. The difference in concepts and trigger words is that we can define relationships and constraints for these concepts making it a more powerful alternative to current implementations. We can continually modify these concepts by simply redefining the knowledge base. If we were to hard-code trigger words, then we would be required to modify the source code of all the systems using these words. In an approach analogous to object oriented software development, we can add new and improved ontology agents without the need to modify any other system.

7.0 Interoperation in Ontology

The Foundation for Intelligent Physical Agents (FIPA) has created a specification for the design of interoperable ontologies [7]. The purpose of developing a FIPA compliant ontology is that it allows for the extendibility of primary ontologies. A FIPA compliant system consists of multiple agents that interact with an Ontology Agent (OA). The Ontology Agent responds to queries for relationships between terms or between ontologies. It could also be used for translating expressions between different ontologies. The Ontology Agent is similar to the mediator in our design. In the FIPA specifications, the Ontology agent accessing the Knowledge Base performs the context analysis. The knowledge base can be accessed via the Open Knowledge Base Connectivity (OKBC) interface. OKBC is used as an interface to connect the front end OA with the back end Knowledge Servers.

The use of OKBC gives the added flexibility of using any existing ontology service such as Ontolingua, Loom or Protege. OKBC also masks the back-end ontology service from the user level agents. Using standard API's in OKBC, the OA can perform all the necessary ontology translations regardless of back-end implementation.

Our design focuses on a FIPA compliant model because of the future benefits

of designing interoperable ontologies. There is also an added advantage of developing ontologies using multiple services.

8.0 Implementation using Beegent and Protégé

The Toshiba Beegent [8] development framework is used for the design of the user-level agents as well as the Ontology Agent. This agent system conforms to the FIPA Agent Communication Language (ACL) specification. The ACL framework in Beegent is encoded in XML allowing for transport via HTTP.

The two main components for the Toshiba Beegent framework are the Agent Wrapper and the Mediation Agent. The Agent Wrapper is used to encapsulate an existing implementation such as the browser functionality for our WWW agent. The Mediation Agent, which is similar to the directory service, processes requests that are passed to it from an Agent wrapper, and takes the necessary action to contact other components in the Beegent framework. The Mediator Agent will also encapsulate functionality to interface with the Knowledge Base (KB).

The KB is developed using Protege. Protege allows for the development of a conceptual ontology using a graphical interface. Fig 1 shows the conceptual design of the diabetes ontology using Protege. Protege includes export capabilities to JDBC allowing for access to the knowledge server via simple database queries. It also supports the inclusion of OKBC framework to access the knowledge server.

9.0 Related Works

There are various studies and projects related to ontologies in healthcare informatics. Some of them are [9]:

- CYC had a project on anatomical terminology
- GALEN is a European community project that aimed to develop a terminology server. It also supports to alignment between various coding systems.
- ONIONS is committed to develop a large scale ontology library for medical ontology and address some of the problems in GALEN
- Snomed-RT is exploiting description logic to develop ontology for clinical environment
- MED is another initiative to develop controlled vocabulary
- Artemis[10] is a research initiative to develop a Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems

10.0 Discussion

Although we have used an interoperable ontology of our telemedicine application, it is unrealistic to assume a global ontology as there are varieties of standards like HL7, CEN TC 251, ISO TC 215, GEHR etc. As more and more ontologies are developed using various standards, there will be a need to merge, align or map between ontologies of the application domain. Unified Medical Language System (UMLS) can be considered as a large merged ontology. More information about merging and alignment can be available from PROMPT[11] , Chimaera [12], FCA-Merge [13] and

SMART [14]. Mapping related information is available in RDFT [15], Knowledge Management Tool [16] and MAFRA [17]. Our future research efforts are concentrated on an agent based framework for automated ontology mapping in healthcare domain.

11.0 References

[1] Howie TC, Kunz JC, Law KH; Software interoperability; Centre of Integrated Facility Engineering; Stanford University; 12/25/1996; http://www.dacs.dtic.mil/techs/interop/title.shtml

[2] Pronab Ganguly & Pradeep Ray; A Methodology for the Development of Software Agent Based Interoperable Telemedicine Systems: A Tele-Electrocardiography Perspective; Telemedicine Journal, Volume 6; Number 2; 2000

[3] Ouksel M Aris; A Framework for Scalable Agent Architecture of Cooperating Heterogenous Knowledge Sources; Intelligent Information Agents; Matthias Klush (Ed.); Springer; 1999; page 100 - 124.

[4] Ramesh v, Canfield Kip, Quirologico Steve & Silva Marcelo; An Intelligent Agent-based Architecture for Interoperability among Heterogeneous Medical Databases; http://hsb.baylor.edu/ramsower/ais.ac.96/papers/ramesh2.htm accessed on 12/23/1999

[5] Fensel, D.; Ontologies: "A silver Bullet for Knowledge Management and Electronic Commerce"; Springer-Verlag Berlin Hiedelberg; 2001.

[6] Peter, J.W., Paul, L.D., Keith, W.T.; "Diabetes and its Management". Blackwell Scientific Publications; 4th edition. 1990

[7] FIPA. FIPA Ontology Service Specification, 2001. http://www.fipa.org/specs/fipa00086/XC00086D.html

[8] Toshiba Corp; Bee-gent WWW page https://www2.toshiba.co.jp/beegent/, accessed on 26/12/00

[9] Aldo Gangemi and Domenico M. Pisanelli and Geri Steve; An Overview of the {ONIONS} Project: Applying Ontologies to the Integration of Medical Terminologies; Journal of "Data Knowledge Engineering", volume 31, number 2, pages 183-220, 1999

[10] Artemis Deliverable D3.1.1.2: Review of the State-of-the Art – Semantic Web and Web Service Semantics

[11]N.F. Noy & M.A Musen; " PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment'; In the Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI – 2000); Austin Tx.

[12] Chimera, http:// www.ksl.stanford.edu/software/chimera

[13] G . Stumme & A Maedche; " (FCA-Merge) Ontology Merging for Federated Ontologies on the Semantic Web"; In Proceedings of the International Workshop for Foundations of Models for Information Integration (FMII-2001), Viterbo, Italy, September 2001

[14] N. F. Noy & M. Musen; "SMART: Automated Support for Ontology Merging and Alignment|; In proceedings of the 12th Workshop on Knowledge Acquisition, Modelling and Management (KAW 1999), Canada, October 1999

[15] B Omelayenko; " Integrating Vocabularies: Discovering and Representing Vocabulary Maps"; In Proceedings of the First International Semantic Web Conference; Sardina, Italy, June 9-12 ; 2002

[16] Crubezy, M, Pincus , Z & Musen M.A; "Mediating Knowledge between Application Components"; In proceedings of the Semantic Integration Workshop of the Second International Semantic Web Conference (ISWC -04) Sanibel Island; Florida

[17]Maedche, A., Motik, D., Silva, N., Volz, R., " MAFRA – A Mapping FRAmework fpr Distributed Ontologies"; In proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management, EKAW -2002, Madrid, Spain

Semantic Interoperability in Telemedicine through Ontology-driven Services

Pronab Ganguly* & Pradeep Ray+
*School of Computer Science and Engineering,
+School of Information Systems & Management
The University of New South Wales, Kensington,
NSW 2052, Australia,
*Tel: 61 2 96915762, Email: pganguly@qantas.com.au
+Tel: 61 2 9385 5890, Email: p.ray@unsw.edu.au

1.0 Introduction

Telemedicine involves the integration of information, human machine and healthcare technologies. As different modalities of patient care require applications running on heterogenous computing environment, software interoperability is a major issue in telemedicine. Software interoperability may be defined as"the capability with which two or more programs can share and process information irrespective of their implementation language and platform" [1]. Our study [2] outlined the development methodology for software agent based interoperable telemedicine systems. Subsequently Ontology has emerged as a major technique for software interoperation. In this paper the importance of ontology will be highlighted from the perspective of semantic interoperation. Section 2.0 describes the issues in software interoperation. Section 3.0 highlights the importance of ontology in software interoperation. Section 4.0, 4.0, 6.0, 7.0 & 8.0 describes the major element of our framework. Section 9.0 describes some related efforts and Section 10.0 describes the scope of further research in this area.

2.0 Software Interoperability

Software interoperability may be defined as the ability for multiple software components to interact regardless of their implementation programming language or hardware platform. The available mechanisms for software interoperability are:

- Physical interoperability: In this approach, the interoperability is achieved by physically transferring the information through electronic media such as floppy disks, CDs and magnetic tapes.
- Data-type interoperability: distributed and disparate programs support structured exchange of information through Application Programming Interfaces (APIs) invoked over a computer network.
- Specification-level interoperability: same as the previous one but also encapsulates knowledge representation differences at the level of abstract data types (e.g. a Table, Tree etc.). This enables programs to communicate at higher levels of abstraction and increases the degree of information hiding. CORBA and Enterprise Java Beans (EJB) fall into this category.

- Semantic interoperability: unlike the above two types of interoperability which are concerned with the form (structured description) at the integration interface, semantic interoperability represents design intent and predicted behaviour as well as form (structured description) of the shared entities. It assumes that different information sources store information on related issues but each may offer a different meaning (semantics) of it.

The semantic interoperability problem may be defined in general as "the ability of a user to access, consistently and coherently, similar (though autonomously defined and managed) classes of digital objects and services distributed across heterogeneous repositories, with federating or mediating software compensating for site-by-site variations"[3].

Thus semantic interoperability between various heterogeneous information sources continues to pose serious challenges to database, artificial intelligence and other related communities. This issue includes system, syntactic and structural/schematic interoperability. When we restrict the context to a scenario where various domain-specific but diverse databases are simultaneously accessed, semantic interoperability may be achieved. However, even in such cases, these databases are heterogeneous in the sense that they store different types of data, different data representation formats and there are different software and computing platforms used to run these databases. Researchers are working towards the objective of identifying the mechanisms that can access data from multiple databases without making changes to the existing databases. Existing approaches for integration of heterogeneous databases include resolution of structural differences between underlying databases. In conceptual or global schema approach, there is a need for standardisation of data structures and definitions [4]. The conceptual schema specifies field and record definitions, structure and rules for updating data values. Various mappings and transformations are used to convert source data into a semantically equivalent and compatible form. The problems with this approach include:

- The emphasis is on schematic (i.e. syntactic) rather than semantic heterogeneity
- It assumes global knowledge is available which is not always possible
- The development costs of coding the system and the data definitions is huge
- The autonomy of individual databases is lost

In a federated database approach, a multidatabase language is used to facilitate interoperability. This language is the vehicle for interoperability. Additionally, each local database provides an export schema (a portion of its overall schema) which it is willing to share with other sources. Each database then uses these export schemas to define an import schema. Thus a partial global schema representing information from other remote databases is generated and used. But the problem is that the users need to determine the meaning of all existing concepts and terms in every database across the network. This limits the number of databases and thus the scalability even in the same domain. It is very difficult to represent semantic information outside a specific domain. Even in a specific domain, human (user) intelligence is usually required to be aware of the context so as to assess semantic information such as in a federated approach. The issues include:

- A specific action can have different results depending on the context. For example, in a telemedicine application, the action "follow a normal diet" will mean different things to different users.

- When using a global schema, it is hard to maintain the autonomy of individual databases and keep the overall system's development costs as low as possible.
- Application-specific solutions that compromise the autonomy, flexibility and scalability of the entire distributed application should be avoided.

Our solution is based on the concepts of *Ontologies* and software agents. The major elements of the proposed system are:

- Since contexts need to be used to represent the underlying semantics of the databases, these contexts are defined in ontology. . So the users only need to express their needs by using terms in the ontology.
- There are *software agents* collaborating with every (human) user in the distributed environment as well as the ontology and the databases. These software agents are considered like human entities (having beliefs, desire and commitments) constrained by pragmatics (intentions, communication etc) as well as semantics (meanings, propositions, validity etc) and syntax (formal structure, data etc). They handle all communication aspects, receive feedback from the user, initiate communication, monitor events and perform certain tasks.

3.0 Ontology and Semantic Interoperation

The word ontology is derived from the Greek words "ontos" (which means "being") and "logos" (which means "word"). Thus ontology is related to various kinds of things that may exist in a given domain. Basically ontology provides a formal specification of the terms in a given domain and relationship among them. Thus Ontology is a formal, explicit specification of a shared conceptualization [5].

- "A 'conceptualization' refers to an abstract model of some scenarios in real life which identifies the relevant concepts of that phenomenon."
- "'Explicit' means that the type of concepts used and the constraints on their use are explicitly defined."
- "'Formal' refers to the fact that the ontology should be computer readable."
- "'Shared' reflects the notion that an ontology captures consensual knowledge, that is, it is not restricted to some individual, but accepted by a group".

A knowledge base consists of an ontology and instances of the associated classes. As the built-in knowledge in ontology is consensual in nature, it can be shared. In the practical sense, development of a domain ontology involves:

- Identification and definition of classes
- Arrangement of the classes in a taxonomical hierarchy
- Attaching properties (slots) and defining allowed values (facets) for these slots
- Allocation of the values for slots for instances

As ontologies are described in a logic based language, the usual features of logic are available. So accurate, distinct and logical linkage as well as distinction can be made among the classes, properties and relations. This helps inference engines o conduct automated reasoning with the help of ontologies. Similarly support is provided to intelligent entities such as software agents. Ontology is a major element for representation of the semantics of the document. Thus ontology can be effectively

used as a mechanism for semantic interoperation in telemedicine as well. In order to illustrate this aspect of ontology, we will first describe a telemedicine application in diabetes management and then highlight importance of ontology as a means for semantic interoperation.

4.0 Telemedicine and Diabetes Management

Diabetes is a condition in which the level of blood glucose is persistently raised above the normal range. If the condition is left untreated it can lead to renal and ocular complications or damage to peripheral nerves [6]. Worldwide, 135 million people suffer from diabetes. The figure is projected to grow by 122% within the next 25 years. These statistics clearly indicate the growing need for an efficient and effective treatment plan for those who suffer from diabetes.

Currently, treatment of diabetes entails limited patient contact with multiple healthcare professionals such as the general practitioner (GP), specialist and clinicians. The interaction would involve regular patient visits to the family GP who collect blood samples to send to the clinical laboratory for testing. The result is reviewed by the GP to determine the patient's condition to prescribe medication. All complicated cases are referred to a specialist.

The above interaction outline the type of organised care currently provided to the diabetic patient. Two major problems identified in the current system are as follows:

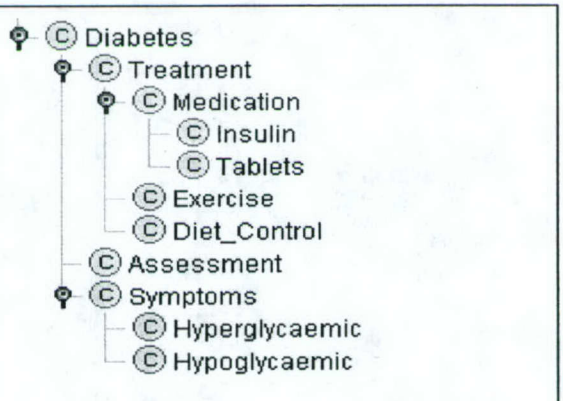


Fig. 1. Simple conceptual definition of the diabetes domain.

- The patient, laboratory clinicians, GP and the specialist interact effectively amongst each other to provide efficient health care to the patient.

- A new approach will be established to maintain consistency within a heterogeneous environment.

We are proposing an ontology-driven multi-agent system for diabetic treatment. This system aims to manage complex human interactions by the use of a diabetes ontology and agent system.
The major components of this implementation are:

- *FIPA Standard based Ontology* development using Protege.
- Toshiba Beegent Platform

The next section presents the need for an ontology-driven agent framework for diabetic management.

5.0 Ontology Driven Software Based Framework

A software agent is a program which performs a specific task on behalf of a user, independently or with little guidance. A software agent also facilitates the communication of programs written in different languages. In the diabetes scenario, each healthcare professional may use different systems to assist in management of patient records and treatment plans. Each system may reside on different platforms across a network. To unify these systems effectively, a software agent could be used. The agent would create a "wrapper" around each application and communicate with other wrappers using an agent communication language. Therefore the healthcare professionals will benefit from an agent-based system to support existing healthcare frameworks.

The behavioural constraints on software agents can be defined so that the agent can change behaviour based on other dynamic components within the system. For example, agent technology can be used to enhance doctor-specialist collaboration. A typical doctor-specialist interaction would involve:

- Doctor contacting specialist about patient
- Specialist formulating treatment plan based on patient information
- Doctor formulating treatment plan based on patient information
- Doctor and Specialist negotiating ideal treatment plan for patient

An agent system can be used to try and emulate part of this behaviour independently of the real doctor or specialist. To effectively emulate this behaviour requires not only the development of an agent, but also the development of ontology. Ontology is a formal, explicit specification of a shared conceptualisation as defined in. It provides a vocabulary of terms and relations with which to model the domain. Ontology is suited to represent high-level information requirements to specify the context information in a collaborative environment.

The ontology will help to share common understanding of the structure of information. Agents can then be used to extract and aggregate information to answer queries to other applications. In the above interaction, if the doctor uses the term "drug" to indicate the medicine that a patient is taking, and the specialist uses the term

"treatment" to indicate the same term, a simple multi-agent framework will not be able to interpret the context of each message although it has the same meaning. To facilitate the conceptual understanding of terms, ontology should be developed. An agent can then query this ontology to perform an intermediate translation of terms for use by other agent systems.

Fig 1 describes a simple conceptual definition for the diabetes domain. Classes (also called concepts) are the main focus of ontologies. Classes are used to describe concepts in the domain. A class can have subclasses that represent concepts more specific than the superclass. In Fig 1 Diabetes represents the main concept containing specific sub concepts such as Treatment and Symptoms. The Treatment concept can be further refined to represent Medication, Exercise, and Diet Control. By defining relationships and constraints for these concepts, we would be able to formally describe the diabetes domain. Once fully developed into a knowledge base, an agent can access these terms and relationships to reason and to derive answers to queries. An ontology driven multi-agent system can then be used to resolve the issue of maintaining consistency with medical terminology and standards.

6.0 Architecture of an Ontology-Driven Agent System

The conceptual model of an ontology-driven agent system is shown in Fig. 2.

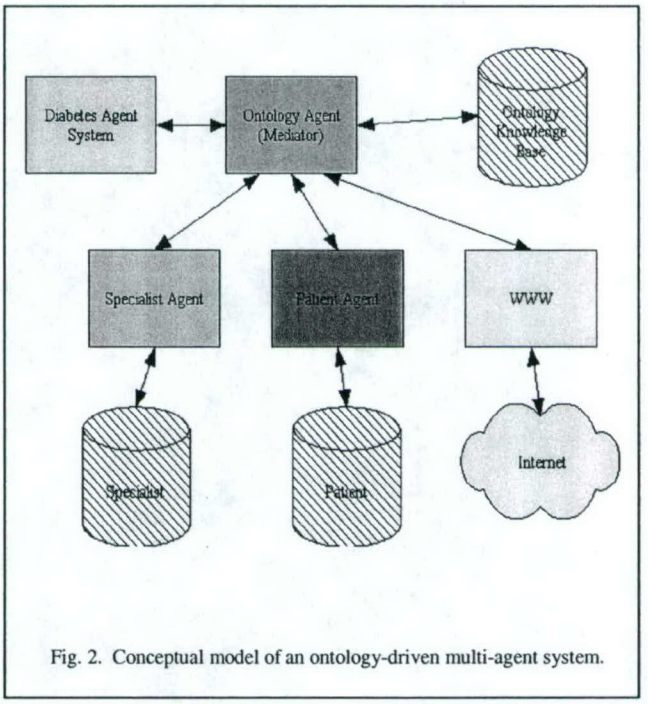


Fig. 2. Conceptual model of an ontology-driven multi-agent system.

The model is based upon an:

- Establishment of agent systems
- Interaction via an agent communication protocol

- Development of an ontology

This model will be in compliance with the Foundation for Intelligent Physical Agents (FIPA) specification explained in the next section.
The following agent systems will be established:

- Diabetes Agent System
- Ontology Agent
- Specialist Agent
- Patient Agent
- WWW Agent

Each agent performs a specific set of function as stated below:

Diabetes Agent System:

- Provides an interface for the GP or specialist. Should incorporate an Eliza based algorithm to respond to the user in natural language.
- Sends any requests from the user to the Ontology Agent (OA).
- Receives messages from the OA to display, or to respond to.

Ontology Agent:

- Receives messages from any agent on the system.
- Messages are broken into tokens to be translated into primary concepts. This interaction takes place in collaboration with the Knowledge Base. For example if a message contains a token called "web" and the primary concept defined in the ontology is "Internet", all messages containing the token web will translated into the token Internet.
- OA then queries all agent systems to determine which agent can satisfy the translated message. The OA then sends the message to the destination agent.

Specialist Agent:

- Provides information about treatment options for a patient. Has access to a knowledge base to determine correct treatment options based on symptoms.
- Can negotiate with Diabetes Agent System to arrive at optimal treatment plan.

Patient Agent:

- Responds to request for patient information. Example, blood glucose levels, symptoms, general patient data
- Should interface with a patient to gather information. Could also incorporate an Eliza based algorithm

WWW Agent:

Provides access to the World Wide Web. Performs a search on data requested by other agents and launches an interface with the requested information.

A scenario of how this system will function is as follows:

The user (GP or specialist) would request for information to the Diabetes Agent System. This request would be sent to the Ontology Agent to be translated based on the Ontology developed. An example request could be to find information from the Internet about diabetes. The ontology agent would traverse each word in the message and translate it into a key concept. For example, the term "web" can be translated into the primary term Internet. The new message would then be: "find information from the Internet about diabetes". This is done so that all messages passed within this system are consistent and can be understood by all agents.

After message translation, the ontology agent has to determine which agent can understand and perform the request. To do so, the ontology agent will query each agent to determine the one that satisfies the key terms in the message. If there is a match then the message is passed to the agent.

For example, the message: "find information on the web about diabetes", would be sent to the WWW Agent. This search agent will then launch a browser to find information on diabetes.

The advantages of developing a system this way is that we are not hard-coding "trigger words" to activate certain actions. Rather we are developing agents to deal with certain concepts within a fixed domain. The difference in concepts and trigger words is that we can define relationships and constraints for these concepts making it a more powerful alternative to current implementations. We can continually modify these concepts by simply redefining the knowledge base. If we were to hard-code trigger words, then we would be required to modify the source code of all the systems using these words. In an approach analogous to object oriented software development, we can add new and improved ontology agents without the need to modify any other system.

7.0 Interoperation in Ontology

The Foundation for Intelligent Physical Agents (FIPA) has created a specification for the design of interoperable ontologies [7]. The purpose of developing a FIPA compliant ontology is that it allows for the extendibility of primary ontologies. A FIPA compliant system consists of multiple agents that interact with an Ontology Agent (OA). The Ontology Agent responds to queries for relationships between terms or between ontologies. It could also be used for translating expressions between different ontologies. The Ontology Agent is similar to the mediator in our design. In the FIPA specifications, the Ontology agent accessing the Knowledge Base performs the context analysis. The knowledge base can be accessed via the Open Knowledge Base Connectivity (OKBC) interface. OKBC is used as an interface to connect the front end OA with the back end Knowledge Servers.

The use of OKBC gives the added flexibility of using any existing ontology service such as Ontolingua, Loom or Protege. OKBC also masks the back-end ontology service from the user level agents. Using standard API's in OKBC, the OA can perform all the necessary ontology translations regardless of back-end implementation.

Our design focuses on a FIPA compliant model because of the future benefits

of designing interoperable ontologies. There is also an added advantage of developing ontologies using multiple services.

8.0 Implementation using Beegent and Protégé

The Toshiba Beegent [8] development framework is used for the design of the user-level agents as well as the Ontology Agent. This agent system conforms to the FIPA Agent Communication Language (ACL) specification. The ACL framework in Beegent is encoded in XML allowing for transport via HTTP.

The two main components for the Toshiba Beegent framework are the Agent Wrapper and the Mediation Agent. The Agent Wrapper is used to encapsulate an existing implementation such as the browser functionality for our WWW agent. The Mediation Agent, which is similar to the directory service, processes requests that are passed to it from an Agent wrapper, and takes the necessary action to contact other components in the Beegent framework. The Mediator Agent will also encapsulate functionality to interface with the Knowledge Base (KB).

The KB is developed using Protege. Protege allows for the development of a conceptual ontology using a graphical interface. Fig 1 shows the conceptual design of the diabetes ontology using Protege. Protege includes export capabilities to JDBC allowing for access to the knowledge server via simple database queries. It also supports the inclusion of OKBC framework to access the knowledge server.

9.0 Related Works

There are various studies and projects related to ontologies in healthcare informatics. Some of them are [9]:

- CYC had a project on anatomical terminology
- GALEN is a European community project that aimed to develop a terminology server. It also supports to alignment between various coding systems.
- ONIONS is committed to develop a large scale ontology library for medical ontology and address some of the problems in GALEN
- Snomed-RT is exploiting description logic to develop ontology for clinical environment
- MED is another initiative to develop controlled vocabulary
- Artemis[10] is a research initiative to develop a Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems

10.0 Discussion

Although we have used an interoperable ontology of our telemedicine application, it is unrealistic to assume a global ontology as there are varieties of standards like HL7, CEN TC 251, ISO TC 215, GEHR etc. As more and more ontologies are developed using various standards, there will be a need to merge, align or map between ontologies of the application domain. Unified Medical Language System (UMLS) can be considered as a large merged ontology. More information about merging and alignment can be available from PROMPT[11] , Chimaera [12], FCA-Merge [13] and

SMART [14]. Mapping related information is available in RDFT [15], Knowledge Management Tool [16] and MAFRA [17]. Our future research efforts are concentrated on an agent based framework for automated ontology mapping in healthcare domain.

11.0 References

[1] Howie TC, Kunz JC, Law KH; Software interoperability; Centre of Integrated Facility Engineering; Stanford University; 12/25/1996; http://www.dacs.dtic.mil/techs/interop/title.shtml

[2] Pronab Ganguly & Pradeep Ray; A Methodology for the Development of Software Agent Based Interoperable Telemedicine Systems: A Tele-Electrocardiography Perspective; Telemedicine Journal, Volume 6; Number 2; 2000

[3] Ouksel M Aris; A Framework for Scalable Agent Architecture of Cooperating Heterogenous Knowledge Sources; Intelligent Information Agents; Matthias Klush (Ed.); Springer; 1999; page 100 - 124.

[4] Ramesh v, Canfield Kip, Quirologico Steve & Silva Marcelo; An Intelligent Agent-based Architecture for Interoperability among Heterogeneous Medical Databases; http://hsb.baylor.edu/ramsower/ais.ac.96/papers/ramesh2.htm accessed on 12/23/1999

[5] Fensel, D.; Ontologies: "A silver Bullet for Knowledge Management and Electronic Commerce"; Springer-Verlag Berlin Hiedelberg; 2001.

[6] Peter, J.W., Paul, L.D., Keith, W.T.; "Diabetes and its Management". Blackwell Scientific Publications; 4th edition. 1990

[7] FIPA. FIPA Ontology Service Specification, 2001. http://www.fipa.org/specs/fipa00086/XC00086D.html

[8] Toshiba Corp; Bee-gent WWW page https://www2.toshiba.co.jp/beegent/, accessed on 26/12/00

[9] Aldo Gangemi and Domenico M. Pisanelli and Geri Steve; An Overview of the {ONIONS} Project: Applying Ontologies to the Integration of Medical Terminologies; Journal of "Data Knowledge Engineering", volume 31, number 2, pages 183-220, 1999

[10] Artemis Deliverable D3.1.1.2: Review of the State-of-the Art – Semantic Web and Web Service Semantics

[11]N.F. Noy & M.A Musen; " PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment'; In the Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI – 2000); Austin Tx.

[12] Chimera, http:// www.ksl.stanford.edu/software/chimera

[13] G . Stumme & A Maedche; " (FCA-Merge) Ontology Merging for Federated Ontologies on the Semantic Web"; In Proceedings of the International Workshop for Foundations of Models for Information Integration (FMII-2001), Viterbo, Italy, September 2001

[14] N. F. Noy & M. Musen; "SMART: Automated Support for Ontology Merging and Alignment|; In proceedings of the 12th Workshop on Knowledge Acquisition, Modelling and Management (KAW 1999), Canada, October 1999

[15] B Omelayenko; " Integrating Vocabularies: Discovering and Representing Vocabulary Maps"; In Proceedings of the First International Semantic Web Conference; Sardina, Italy, June 9-12 ; 2002

[16] Crubezy, M, Pincus , Z & Musen M.A; "Mediating Knowledge between Application Components"; In proceedings of the Semantic Integration Workshop of the Second International Semantic Web Conference (ISWC -04) Sanibel Island; Florida

[17]Maedche, A., Motik, D., Silva, N., Volz, R., " MAFRA – A Mapping FRAmework fpr Distributed Ontologies"; In proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management, EKAW -2002, Madrid, Spain